

High-dimensional integration on \mathbb{R}^d , weighted Hermite spaces, and orthogonal transforms

Christian Irrgeher* and Gunther Leobacher†

September 23, 2014

Abstract

It has been found empirically that quasi-Monte Carlo methods are often efficient for very high-dimensional problems, that is, with dimension in the hundreds or even thousands. The common explanation for this surprising fact is that those functions for which this holds true behave rather like low-dimensional functions in that only few of the coordinates have a sizeable influence on its value. However, this statement may be true only after applying a suitable orthogonal transform to the input data, like utilizing the Brownian bridge construction or principal component analysis construction.

We study the effect of general orthogonal transforms on functions on \mathbb{R}^d which are elements of certain weighted reproducing kernel Hilbert spaces. The notion of Hermite spaces is defined and it is shown that there are examples which admit tractability of integration. We translate the action of the orthogonal transform of \mathbb{R}^d into an action on the Hermite coefficients and we give examples where orthogonal transforms have a dramatic effect on the weighted norm, thus providing an explanation for the efficiency of using suitable orthogonal transforms.

Keywords: Quasi-Monte Carlo, tractability, effective dimension, orthogonal transforms

MSC2010: 65D30, 65Y20

1 Introduction

Many important problems from quantitative finance and other applications of probability theory can be written as expected values of functions depending on the path of a Brownian motion, i.e.

$$\mathbb{E}(g(B)) = ? \tag{1}$$

where $B = (B_t)_{t \in [0, T]}$ is an m -dimensional standard Brownian motion with index set $[0, T]$ and g is a function such that the expected value is well-defined. The most prominent example from finance is the celebrated Black-Scholes equation for option pricing. In its general form it

*Christian Irrgeher, Institute of Financial Mathematics, Johannes Kepler University Linz, Altenbergerstraße 69, A-4040 Linz, Austria. e-mail: christian.irrgeher@jku.at

†Gunther Leobacher, Institute of Financial Mathematics, Johannes Kepler University Linz, Altenbergerstraße 69, A-4040 Linz, Austria. e-mail: gunther.leobacher@jku.at.

states that the problem of determining the value of a European-style contingent claim – like a European or Asian option – in an arbitrage-free market model driven by a Brownian motion, can be formulated as the computation of an expected value of the form (1), see e.g. Björk [3].

In most cases of practical interest these problems cannot be solved in closed form, and thus it is necessary to approximate (1) numerically. A very versatile way of doing this is by using Monte Carlo, that is, a large number of random paths of the Brownian motion is generated and the function value of g over this sample is averaged. Usually, the actual Monte Carlo simulation requires the discretization of the Brownian path. One thus uses a randomized algorithm $Q_{n,d}$, where d denotes the dimension of the discretized problem and n corresponds to the number of (discretized) paths over which the average is taken.

Also if a deterministic algorithm, like quasi-Monte Carlo, is to be applied, the problem has first to be discretized. For that, we use a time grid of d_0 equidistant points to discretize the time interval $[0, T]$ such that we get a discrete m -dimensional Brownian path $(B_{1/d_0}, B_{2/d_0}, \dots, B_{d_0/d_0})$ with $B_{i/d_0} \in \mathbb{R}^m$ for $i = 1, \dots, d_0$. Moreover, we have to choose a $d = m \times d_0$ -dimensional function g_{d_0} depending on the discrete Brownian path $(B_{1/d_0}, B_{2/d_0}, \dots, B_{d_0/d_0})$ such that $\mathbb{E}(g(B)) \approx \mathbb{E}(g_{d_0}(B_{1/d_0}, B_{2/d_0}, \dots, B_{d_0/d_0}))$. So the problem is approximated by the multivariate integration problem $\mathbb{E}(g_{d_0}(B_{1/d_0}, B_{2/d_0}, \dots, B_{d_0/d_0}))$ of dimension d which can be solved, e.g., by quasi-Monte Carlo integration, see Dick and Pillichshammer [7]. Note that, here the integration problem for some function f is given over \mathbb{R}^d and thus, a quasi-Monte Carlo method for computing this integral is given by an equally weighted quadrature rule $Q_{n,d}(f, \mathcal{P}) = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i)$ with deterministic point set $\mathcal{P} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^d$.

In contrast to Monte Carlo, the method with which the discrete Brownian paths are generated makes a difference for quasi-Monte Carlo methods. There are three classical Brownian path construction methods: the forward method (or step-by-step method), the Brownian bridge construction (a.k.a. Lévy-Ciesielski construction) and the principal component analysis (PCA) construction. In Moskowitz and Caflisch [14] the convergence of QMC integration is dramatically improved for some examples from finance by using the Brownian Bridge construction. Similar results are presented by Acworth et al. [1] for the PCA construction. However, Papageorgiou [15] shows that there are problems for which the forward construction gives faster convergence than the Brownian bridge construction. He further shows that every linear construction method for a one-dimensional Brownian path corresponds to a unique orthogonal transform of the \mathbb{R}^d . Subsequently, he formulates an equivalence principle that roughly states that every construction that is good for one problem is bad for another problem. Wang and Sloan [21] give a more general version of that equivalence principle.

Due to the equivalence principle, the choice of the orthogonal transform should depend on the integration problem, as given by the function g . Leobacher [13], proposes to restrict this search to transforms which can be applied using $\mathcal{O}(d \log(d))$ operations. In Imai and Tan [9] as well as in Irrgeher and Leobacher [10, 11] methods are presented to find, for a given problem, an (in some sense) optimal orthogonal transform. In both approaches the integration problem is linearized and the “optimal” transform is determined for the linear problem. Imai and Tan use a Taylor expansion of order 1 for the linearization step whereas in Irrgeher and Leobacher [10] a linear regression is performed. The idea behind both methods is to minimize the effective

dimension, as defined by Caffisch, Morokoff, and Owen [5].

While the concept of effective dimension has its merits in explaining the effectiveness of deterministic methods for high-dimensional problems, it also does have some drawbacks. One of those drawbacks is that it does not interact smoothly with orthogonal transforms. For example, consider some continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$ with $\int_{\mathbb{R}} f(x)^2 \varphi(x) dx < \infty$ and $\int_{\mathbb{R}} f(x) \varphi(x) dx = 0$, where $\varphi(x) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{x^2}{2})$ is the standard normal density. Then the function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $g(x_1, x_2) := f(x_1)$ is continuous with $\int_{\mathbb{R}^2} g(x_1, x_2)^2 \varphi(x_1) \varphi(x_2) dx_1 dx_2 < \infty$ and $\int_{\mathbb{R}^2} g(x_1, x_2) \varphi(x_1) \varphi(x_2) dx_1 dx_2 = 0$. The ANOVA decomposition of g is given by $g_{\{1\}} = f$, $g_{\emptyset} = g_{\{2\}} = g_{\{1,2\}} \equiv 0$. On the other hand, consider $\tilde{g} = g \circ U$, where U is some orthogonal transform of the \mathbb{R}^2 . Then $\tilde{g}(x_1, x_2) = f(cx_1 + sx_2)$ with $c^2 + s^2 = 1$, and therefore we have $\tilde{g}_{\{1,2\}} = \tilde{g}$, and $\tilde{g}_{\emptyset} = \tilde{g}_{\{1\}} = \tilde{g}_{\{2\}} \equiv 0$, for every choice of c, s with $|c| \neq 1$ and $|s| \neq 1$. Thus even the slightest rotation transforms a function of effective dimension 1 into one with effective dimension 2 (both in the truncation and in the superposition sense). It is not hard to construct multidimensional examples where a slight orthogonal transform effects an arbitrary change in effective dimension.

We have a competing explanation for the efficiency of QMC, namely the theory of weighted spaces as proposed by Sloan and Woźniakowski [18]. They introduce weighted norms on Sobolev spaces that assign different degrees of importance to different coordinates. The idea is related to the concept of effective dimension in that both concepts concentrate on problems for which only relatively few input parameters are really important. From the point of view of orthogonal transforms it is now interesting to ask whether one can concatenate the original function with an orthogonal transform in a way that makes the weighted norm of the resulting function – and thus the integration error – small. Before we proceed with this program, we need to go back one step.

We are interested in analyzing the error which occurs by approximating the expected value $\mathbb{E}(g(B))$ by using a QMC algorithm $Q_{n,d}$, where d gives the dimension of the discretized space and n is the number of integration nodes. The error is given by

$$\text{err} := |\mathbb{E}(g(B)) - Q_{n,d}|.$$

As mentioned earlier, the function g has to be approximated by a $d = m \times d_0$ -dimensional function g_{d_0} which depends on a discrete m -dimensional Brownian path. Then there further exists a function $f_d : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $f_d(X) = g_{d_0}((B_{T/d_0}, B_{2T/d_0}, \dots, B_T))$ where $X = (X_1, \dots, X_d)$ is a standard Gaussian vector. Hence,

$$\begin{aligned} \text{err} &= |\mathbb{E}(g(B)) - \mathbb{E}(f_d(X)) + \mathbb{E}(f_d(X)) - Q_{n,d}| \\ &\leq \underbrace{|\mathbb{E}(g(B)) - \mathbb{E}(f_d(X))|}_{=: \text{err}_{\text{disc}}} + \underbrace{|\mathbb{E}(f_d(X)) - Q_{n,d}|}_{=: \text{err}_{\text{int}}}. \end{aligned} \quad (2)$$

err_{disc} is called the *discretization error* and err_{int} is called the *integration error*. That means, we can bound the total error by the sum of the error coming from the discretization and the error which comes from the QMC integration. Now it is obvious that we prefer to approximate

the function g by a function g_{d_0} or f_d , respectively, such that the discretization error err_{disc} becomes small. But it is equally important that err_{int} does not explode with growing d .

An important practical example is the case where $g(B)$ is described by a solution of a stochastic differential equation (SDE), e.g., $g(B) = \psi(X_T)$ where X is a stochastic process that solves

$$\begin{aligned} X_0 &= x_0 \\ dX_t &= a(t, X_t)dt + b(t, X_t)dB_t \end{aligned}$$

where a is a \mathbb{R}^k -valued function and b is a $\mathbb{R}^{k \times m}$ -valued function. There are many different discretization methods in the theory of stochastic differential equation which can be applied to $g(B)$. The two best-known are the Euler-Maruyama method and the Milstein method, but there are also higher-order Runge-Kutta methods. For more information about these discretization methods we refer to Kloeden and Platen [12] which also provides an extensive analysis of the convergence of these methods. In [12] it is shown that, under mild conditions on the coefficient functions a and b , as well as the “payoff function” ψ , we have an upper bound on the discretization error of the form

$$\text{err}_{\text{disc}} = |\mathbb{E}(g(B)) - \mathbb{E}(f_d(X))| \leq cd^{-q}, \quad (3)$$

where $c > 0$ is constant and $q > 0$ is the convergence rate which depends on the discretization method used and on the function ψ . In Glasserman [8, Chapter 6] it is discussed how these discretization methods are applied to problems coming from finance. So for this kind of problems (3) provides us with an estimate on err_{disc} in inequality (2).

The emphasis of this paper lies on the analysis of the integration error err_{int} to get an upper bound on the total error. Moreover, we are interested in the behavior of the integration error with respect to d , because we know that the discretization error can only be reduced by increasing the dimension d . This will lead us to the study of QMC-tractability. In particular, we want to know how orthogonal transforms affect the weighted norm and whether the growth of the complexity of integration can be moderated by the use of suitable orthogonal transforms. To that end we consider special spaces of integrable functions on the \mathbb{R}^d which we call *Hermite spaces*. Those spaces are spanned by Hermite polynomials, which enjoy a certain invariance under orthogonal transforms. It turns out that this class of spaces contains examples of weighted spaces for which tractability can be proven and which are sufficiently rich to contain interesting functions.

The remainder of the paper is organized in the following way. We recall the basic facts about Gaussian measures and Hermite polynomials in Section 2 and we discuss the Hermite expansion of functions. In one of the main parts of the paper, namely in Section 3, we introduce Hermite spaces, present basic properties and discuss multivariate integration in these spaces. Furthermore, we investigate the tractability of QMC methods in these spaces and we discuss what kind of functions belong to Hermite spaces. In the second main part, Section 4, we show that orthogonal transformations can be used to reduce the weighted norm of a function, thus improving the convergence of a QMC algorithm. We further give a representation of the operator

\mathcal{A}_U on the Hermite space that is induced by an orthogonal transform U of the \mathbb{R}^d in terms of direct sums of tensor powers of U . Finally, Section 5 summarizes the most important results and states some open problems.

2 Hermite polynomials and Hermite expansion

In this section we recall some basic definitions and results concerning Gaussian measure, the Hermite polynomials and the Hermite expansion.

2.1 Gaussian measure

Definition 2.1. The *standard Gaussian measure* is a Borel probability measure on the \mathbb{R}^d with density $\varphi_d : \mathbb{R}^d \rightarrow \mathbb{R}$ given by

$$\varphi_d(\mathbf{x}) = (2\pi)^{-\frac{d}{2}} e^{-\frac{\mathbf{x}^\top \mathbf{x}}{2}}$$

with respect to the d -dimensional Lebesgue measure.

The notation $\mathbf{x}^\top \mathbf{y}$, as it is used in the above definition, denotes the inner product in the Euclidean space, i.e., for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\mathbf{x}^\top \mathbf{y} = \sum_{k=1}^d x_k y_k$ with $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{y} = (y_1, \dots, y_d)$. We will exclusively work with the standard Gaussian measure. A more general discussion about Gaussian measures can be found in Bogachev [4]. Furthermore, we will write φ instead of φ_1 in the one-dimensional case.

We say that a measurable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is Gaussian square-integrable if

$$\int_{\mathbb{R}^d} f(\mathbf{x})^2 \varphi_d(\mathbf{x}) d\mathbf{x} < \infty,$$

and we denote the linear space of Gaussian square-integrable functions by $\mathcal{L}^2(\mathbb{R}^d, \varphi_d)$.

Moreover, we say that two functions f and g are equivalent if $f = g$ almost everywhere. Then, the space of all equivalence classes of Gaussian square-integrable functions on the \mathbb{R}^d , denoted by $L^2(\mathbb{R}^d, \varphi_d)$, forms a Hilbert space with inner product

$$\langle f, g \rangle_{L^2(\mathbb{R}^d, \varphi_d)} = \int_{\mathbb{R}^d} f(\mathbf{x}) g(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x}.$$

The corresponding norm is induced by the inner product, i.e., $\|f\|_{L^2(\mathbb{R}^d, \varphi_d)} = \sqrt{\langle f, f \rangle_{L^2(\mathbb{R}^d, \varphi_d)}}$.

2.2 Hermite polynomials

Before we introduce the multivariate Hermite polynomials, we recall multi-index notation: a d -dimensional *multi-index* is a d -tuple $\mathbf{k} = (k_1, \dots, k_d)$ with nonnegative integer entries. We use the following conventions for $\mathbf{k}, \mathbf{j} \in \mathbb{N}_0^d$ and $\mathbf{x} \in \mathbb{R}^d$:

- (i) $\mathbf{k} \leq \mathbf{j} \iff k_i \leq j_i \text{ for all } i = 1, \dots, d;$
- (ii) $\mathbf{k} + \mathbf{j} = (k_1 + j_1, \dots, k_d + j_d);$

$$(iii) \quad |\mathbf{k}| = k_1 + \dots + k_d;$$

$$(iv) \quad \mathbf{k}! = k_1! \dots k_d!;$$

$$(v) \quad \mathbf{x}^{\mathbf{k}} = x_1^{k_1} \dots x_d^{k_d}.$$

There are some slightly different definitions of Hermite polynomials in the literature which are all related, see, e.g., Bogachev [4], Sansone [17] or Thangavelu [20]. In this paper we use the definition of Hermite polynomials such that they are the Gram-Schmidt orthonormalization of the polynomials $1, x, x^2, x^3, \dots$ with respect to the standard Gaussian measure. This corresponds to the definition used in [4].

Definition 2.2. For any $k \in \mathbb{N}_0$ we call

$$H_k(x) = \frac{(-1)^k}{\sqrt{k!}} e^{\frac{x^2}{2}} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}}$$

the k -th (univariate) Hermite polynomial and for any multi-index $\mathbf{k} \in \mathbb{N}_0^d$ the (multivariate) Hermite polynomial $H_{\mathbf{k}}(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^d$ is defined by

$$H_{\mathbf{k}}(\mathbf{x}) = \prod_{j=1}^d H_{k_j}(x_j).$$

The exponential generating function G of the Hermite polynomials is given by

$$G(\mathbf{x}, \mathbf{t}) := e^{\mathbf{x}^\top \mathbf{t} - \frac{\mathbf{t}^\top \mathbf{t}}{2}} = \sum_{\mathbf{k} \in \mathbb{N}_0^d} H_{\mathbf{k}}(\mathbf{x}) \frac{\mathbf{t}^{\mathbf{k}}}{\sqrt{\mathbf{k}!}} \quad (4)$$

with $\mathbf{x}, \mathbf{t} \in \mathbb{R}^d$, see Bogachev [4, Chapter 1.3] for the case of $d = 1$. The exponential generating function and the representation of the Hermite polynomials via function G , i.e., $H_{\mathbf{k}}(\mathbf{x}) = \sqrt{\mathbf{k}!} \frac{d^{\mathbf{k}}}{d\mathbf{t}^{\mathbf{k}}} G(\mathbf{x}, \mathbf{t})|_{\mathbf{t}=\mathbf{0}}$, is a useful tool to work with these kind of polynomials.

Proposition 2.3. The sequence of Hermite polynomials $(H_{\mathbf{k}}(\mathbf{x}))_{\mathbf{k} \in \mathbb{N}_0^d}$ forms an orthonormal basis of the function space $L^2(\mathbb{R}^d, \varphi_d)$.

Proof. A proof can be found in Bogachev [4], Lemma 1.3.2. and Corollary 1.3.3. \square

Let $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{N}_0^d$ with the 1 at the i -th entry. Then for $\mathbf{k}, \ell \in \mathbb{N}_0^d$ and for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have the following properties:

- (i) $\left| H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x})^{\frac{1}{2}} \right| \leq \mu^d (2\pi)^{-\frac{d}{4}} \leq 1$ with constant μ smaller than 1.086435 (Cramer's bound);
- (ii) $H_{\mathbf{k}+\mathbf{e}_i}(\mathbf{x}) = \frac{1}{\sqrt{k_i+1}} \left(x_i H_{\mathbf{k}}(\mathbf{x}) - \frac{\partial}{\partial x_i} H_{\mathbf{k}}(\mathbf{x}) \right);$
- (iii) $\frac{\partial^{|\ell|}}{\partial \mathbf{x}^\ell} H_{\mathbf{k}}(\mathbf{x}) = \begin{cases} \sqrt{\frac{\mathbf{k}!}{(\mathbf{k}-\ell)!}} H_{\mathbf{k}-\ell}(\mathbf{x}) & \text{if } \mathbf{k} \geq \ell; \\ 0 & \text{otherwise.} \end{cases}$

This formulation of Cramer's bound follows from the slightly different statement for an alternative definition of the Hermite polynomials in Sansone [17] Section 4.5, where the author refers to Cramer [6]. The properties (ii) and (iii) are easily deduced from the Definition 2.2 and (4). Next we define a first order differential operator \mathcal{D}_x as

$$\mathcal{D}_x := \frac{\partial}{\partial x} - x, \quad (5)$$

which is motivated by the above Property (ii). Moreover, we have that $-\mathcal{D}_{x_i}$ is the adjoint operator of $\frac{\partial}{\partial x_i}$ with respect to the $L^2(\mathbb{R}^d, \varphi_d)$ inner product,

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{\partial}{\partial x_i} (H_{\mathbf{k}}(\mathbf{x})) H_{\ell}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} &= - \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x}) \frac{\partial}{\partial x_i} (H_{\ell}(\mathbf{x}) \varphi_d(\mathbf{x})) d\mathbf{x} \\ &= - \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x}) \left(\frac{\partial}{\partial x_i} H_{\ell}(\mathbf{x}) - x_i H_{\ell}(\mathbf{x}) \right) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= - \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x}) \mathcal{D}_{x_i} (H_{\ell}(\mathbf{x})) \varphi_d(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Recall that we have for the differential operator $\mathcal{D} = (\mathcal{D}_{x_1}, \dots, \mathcal{D}_{x_d})$ the multi-index notation $\mathcal{D}^{\ell} = \mathcal{D}_{x_1}^{\ell_1} \dots \mathcal{D}_{x_d}^{\ell_d}$ for any $\ell \in \mathbb{N}_0^d$.

2.3 Hermite expansion

Since we know from Proposition 2.3 that the Hermite polynomials form an orthonormal basis of $L^2(\mathbb{R}^d, \varphi_d)$, we can write any function $f \in L^2(\mathbb{R}^d, \varphi_d)$ as a Hermite series, i.e.,

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) \quad (6)$$

where the sum converges in the $L^2(\mathbb{R}^d, \varphi_d)$ -norm and $\hat{f}(\mathbf{k})$ is the \mathbf{k} -th Hermite coefficient given by

$$\hat{f}(\mathbf{k}) = \int_{\mathbb{R}^d} f(\mathbf{x}) H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x}.$$

Note that the \mathbf{k} -th Hermite coefficient exists for every $f \in L^2(\mathbb{R}^d, \varphi_d)$, because

$$\int_{\mathbb{R}^d} |f(\mathbf{x}) H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x})| d\mathbf{x} \leq \left(\int_{\mathbb{R}^d} f(\mathbf{x})^2 \varphi_d(\mathbf{x}) d\mathbf{x} \right)^{1/2} \left(\int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x})^2 \varphi_d(\mathbf{x}) d\mathbf{x} \right)^{1/2} < \infty.$$

Furthermore we can show that the Hermite expansion is unique under the assumption of continuity.

Proposition 2.4. *Let $f \in L^2(\mathbb{R}^d, \varphi_d)$ be a continuous function. If $\hat{f}(\mathbf{k}) = 0$ for all $\mathbf{k} \in \mathbb{N}_0^d$, then $f \equiv 0$.*

Proof. We know that $f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x})$ holds in the $L^2(\mathbb{R}^d, \varphi_d)$ sense and thus, using

the Parseval identity

$$\int_{\mathbb{R}^d} f(\mathbf{x})^2 \varphi_d(\mathbf{x}) d\mathbf{x} = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k})^2 = 0.$$

Since f is continuous, we get that $f \equiv 0$. \square

Corollary 2.5. *Let $f, g \in L^2(\mathbb{R}^d, \varphi_d)$ be continuous functions with the same Hermite coefficients. Then $f = g$.*

Proof. The statement follows by applying Proposition 2.4 to $f - g$. \square

In general, the Hermite expansion (6) of an $L^2(\mathbb{R}^d, \varphi_d)$ -function does not need to converge for fixed $\mathbf{x} \in \mathbb{R}^d$, but under further assumptions on the function as well as on its Hermite coefficients the Hermite expansion converges even pointwise.

Proposition 2.6. *Let $f \in L^2(\mathbb{R}^d, \varphi_d)$ be continuous with $\sum_{\mathbf{k} \in \mathbb{N}_0^d} |\hat{f}(\mathbf{k})| < \infty$. Then*

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^d$.

Proof. Because of Cramer's bound, see Property (i) of the Hermite polynomials, we get

$$|\hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x})^{\frac{1}{2}}| \leq |\hat{f}(\mathbf{k})|$$

for each $\mathbf{k} \in \mathbb{N}_0^d$. Since $\sum_{\mathbf{k} \in \mathbb{N}_0^d} |\hat{f}(\mathbf{k})| < \infty$, Weierstrass' uniform convergence theorem, see e.g. Rudin [16] Theorem 7.10, states that $\sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x})^{\frac{1}{2}}$ converges uniformly towards a function \bar{f} . Furthermore, due to the uniform limit theorem, see e.g. Rudin [16] Theorem 7.12, we have that \bar{f} is continuous, because $H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x})^{\frac{1}{2}}$ is continuous for each $\mathbf{k} \in \mathbb{N}_0^d$. Thus, we even get that

$$\sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) = \bar{f}(\mathbf{x}) \varphi_d(\mathbf{x})^{-\frac{1}{2}}$$

holds for all $\mathbf{x} \in \mathbb{R}^d$. Since $\mathbf{x} \mapsto \bar{f}(\mathbf{x}) \varphi_d(\mathbf{x})^{-\frac{1}{2}}$ has the same Hermite coefficients as f and both functions are continuous, we know from the uniqueness of the Hermite expansion that $f(\mathbf{x}) = \bar{f}(\mathbf{x}) \varphi_d(\mathbf{x})^{-\frac{1}{2}}$ for all $\mathbf{x} \in \mathbb{R}^d$. So we end up with

$$\sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) = f(\mathbf{x}),$$

which holds for all $\mathbf{x} \in \mathbb{R}^d$. \square

3 Quasi-Monte Carlo integration in weighted Hermite spaces

In this section we study quasi-Monte Carlo integration with respect to the Gaussian measure. For that, we set up a reproducing kernel Hilbert space of functions on the \mathbb{R}^d and we further show that under additional assumptions QMC integration in these spaces is polynomially tractable.

3.1 Hermite spaces

Definition 3.1. Let $\mathbf{r} : \mathbb{N}_0^d \rightarrow \mathbb{R}^+$ be a summable function, i.e., $\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) < \infty$. For $f \in L^2(\mathbb{R}^d, \varphi_d)$ let

$$\|f\|_{\mathbf{r}} := \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k})^{-1} |\hat{f}(\mathbf{k})|^2 \right)^{\frac{1}{2}}$$

where $\hat{f}(\mathbf{k})$ is the \mathbf{k} -th Hermite coefficient of f . Then we call

$$\mathcal{H}_{\mathbf{r}} := \left\{ f \in \mathcal{L}^2(\mathbb{R}^d, \varphi_d) \cap C(\mathbb{R}^d) : \|f\|_{\mathbf{r}} < \infty \right\}$$

a *Hermite space*.

Note that $\|\cdot\|_{\mathbf{r}}$ is a semi-norm on $\mathcal{L}^2(\mathbb{R}^d, \varphi_d)$, but under the additional assumption, that the functions are continuous as well, it is a norm. Thus, a Hermite space is a Banach space with norm $\|\cdot\|_{\mathbf{r}}$; in fact it is a Hilbert space when equipped with the inner product

$$\begin{aligned} \langle f, g \rangle_{\mathbf{r}} &:= \frac{1}{4} \left(\|f + g\|_{\mathbf{r}}^2 - \|f - g\|_{\mathbf{r}}^2 \right) \\ &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k})^{-1} \hat{f}(\mathbf{k}) \hat{g}(\mathbf{k}). \end{aligned}$$

Theorem 3.2. Let $f \in \mathcal{H}_{\mathbf{r}}$. Then, the Hermite expansion of f converges pointwise, i.e.,

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x})$$

holds for all $\mathbf{x} \in \mathbb{R}^d$.

Proof. We have that f is a Gaussian square-integrable and continuous function. Moreover, we get from the Cauchy-Schwarz inequality

$$\begin{aligned} \sum_{\mathbf{k} \in \mathbb{N}_0^d} |\hat{f}(\mathbf{k})| &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} |\mathbf{r}(\mathbf{k})^{\frac{1}{2}} \mathbf{r}(\mathbf{k})^{-\frac{1}{2}} \hat{f}(\mathbf{k})| \\ &\leq \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) \right)^{\frac{1}{2}} \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k})^{-1} \hat{f}(\mathbf{k})^2 \right)^{\frac{1}{2}} \\ &= \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) \right)^{\frac{1}{2}} \|f\|_{\mathbf{r}} < \infty, \end{aligned}$$

because \mathbf{r} is summable. The statement now follows from Proposition 2.6. \square

We shall show shortly that a Hermite space is always a reproducing kernel Hilbert space. For that, we first recall the definition of a reproducing kernel Hilbert space as well as some important properties.

Definition 3.3. A Hilbert space \mathcal{H} of functions $f : X \rightarrow \mathbb{R}$ on a set X with inner product $\langle \cdot, \cdot \rangle$ is called a reproducing kernel Hilbert space, if there exists a function $K : X \times X \rightarrow \mathbb{R}$ such that

- (i) $K(\cdot, y) \in \mathcal{H}$ for each fixed $y \in X$ and
- (ii) $\langle f, K(\cdot, y) \rangle = f(y)$ for each fixed $y \in X$ and for all $f \in \mathcal{H}$ (reproducing property).

In this definition we use the notation $K(\cdot, y)$ to indicate that K is a function of the first variable, i.e., $x \mapsto K(x, y)$, and the inner product $\langle f, K(\cdot, y) \rangle$ is taken with respect to the first variable of K . The function K is called reproducing kernel and is unique: Assume there is another function $\tilde{K}(\cdot, y) \in \mathcal{H}$ satisfying the reproducing property. Then, with the reproducing property of K and \tilde{K} we get

$$\tilde{K}(x, y) = \langle \tilde{K}(\cdot, y), K(\cdot, x) \rangle = \langle K(\cdot, x), \tilde{K}(\cdot, y) \rangle = K(y, x) = K(x, y)$$

where the last equality holds because of the symmetry of the reproducing kernel,

$$K(x, y) = \langle K(\cdot, y), K(\cdot, x) \rangle = \langle K(\cdot, x), K(\cdot, y) \rangle = K(y, x).$$

More information about the theory of reproducing kernel Hilbert spaces can be found in Aronszajn [2].

Theorem 3.4. A Hermite space $\mathcal{H}_{\mathbf{r}}$ is a reproducing kernel Hilbert space and the reproducing kernel function $K_{\mathbf{r}} : \mathbb{R}^d \times \mathbb{R}^d$ is given by

$$K_{\mathbf{r}}(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) H_{\mathbf{k}}(\mathbf{y}). \quad (7)$$

Proof. First, we show that $K_{\mathbf{r}}(\cdot, \mathbf{y})$ belongs to the Hermite space $\mathcal{H}_{\mathbf{r}}$ for each $\mathbf{y} \in \mathbb{R}^d$. The \mathbf{k} -th Hermite coefficient of $K_{\mathbf{r}}$ as a function in \mathbf{x} is

$$\begin{aligned} \widehat{K_{\mathbf{r}}(\cdot, \mathbf{y})}(\mathbf{k}) &= \int_{\mathbb{R}^d} K_{\mathbf{r}}(\mathbf{x}, \mathbf{y}) H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= \sum_{\mathbf{j} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{j}) \int_{\mathbb{R}^d} H_{\mathbf{j}}(\mathbf{x}) H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} H_{\mathbf{j}}(\mathbf{y}) \\ &= \mathbf{r}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{y}). \end{aligned}$$

So we get that

$$\|K_{\mathbf{r}}(\cdot, \mathbf{y})\|_{\mathbf{r}}^2 = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k})^{-1} \left(\widehat{K_{\mathbf{r}}(\cdot, \mathbf{y})}(\mathbf{k}) \right)^2$$

$$\begin{aligned}
&= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k})^{-1} \mathbf{r}(\mathbf{k})^2 H_{\mathbf{k}}(\mathbf{y})^2 \\
&\leq c e^{\frac{\mathbf{y}^\top \mathbf{y}}{2}} \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) < \infty,
\end{aligned}$$

where Cramer's bound is used to estimate $H_{\mathbf{k}}(\mathbf{y})$ by $c e^{\frac{\mathbf{y}^\top \mathbf{y}}{2}}$ with positive constant c . Next we have to verify that $K_{\mathbf{r}}$ satisfies the reproducing property. For any $f \in \mathcal{H}_{\mathbf{r}}$ we have that

$$\begin{aligned}
\langle f, K_{\mathbf{r}}(\cdot, \mathbf{y}) \rangle_{\mathbf{r}} &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k})^{-1} \hat{f}(\mathbf{k}) \widehat{K_{\mathbf{r}}(\cdot, \mathbf{y})}(\mathbf{k}) \\
&= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k})^{-1} \hat{f}(\mathbf{k}) \mathbf{r}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{y}) \\
&= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{y}) \\
&= f(\mathbf{y})
\end{aligned}$$

holds. Thus, the function $K_{\mathbf{r}}$, given by (7), is indeed the reproducing kernel of the Hermite space $\mathcal{H}_{\mathbf{r}}$ and consequently, $\mathcal{H}_{\mathbf{r}}$ is a reproducing kernel Hilbert space. \square

It is well-known that function evaluation in a reproducing kernel Hilbert space is a continuous linear functional on that space.

We give two examples of Hermite spaces where the function \mathbf{r} controls both the decay of the Hermite coefficients of the functions and the influence of the coordinates of the variables.

3.1.1 Weighted Hermite spaces with polynomially decaying coefficients

Let $\alpha > 1$ and $\gamma > 0$. Then we define a function $p_{\alpha, \gamma} : \mathbb{N}_0 \rightarrow \mathbb{R}^+$ as

$$p_{\alpha, \gamma}(k) := \begin{cases} 1 & \text{if } k = 0, \\ \gamma k^{-\alpha} & \text{if } k \neq 0 \end{cases}$$

and we note that

$$\sum_{k=0}^{\infty} p_{\alpha, \gamma}(k) = 1 + \gamma \sum_{k=1}^{\infty} k^{-\alpha} = 1 + \gamma \zeta(\alpha) < \infty$$

where ζ is the Riemann zeta function. Since $p_{\alpha, \gamma}$ is summable, we can set $\mathbf{r} = p_{\alpha, \gamma}$ and consider the Hermite space $\mathcal{H}_{p_{\alpha, \gamma}}$ of functions on the \mathbb{R} . Then, the norm of $\mathcal{H}_{p_{\alpha, \gamma}}$ can be written as

$$\|f\|_{p_{\alpha, \gamma}}^2 = 1 + \gamma^{-1} \sum_{k=1}^{\infty} k^{\alpha} \hat{f}(k)^2.$$

Hence, the Hermite coefficients of f have to decay polynomially with a rate of $o(k^{-\alpha/2})$ so that the norm becomes finite and, consequently, the function f belongs to the Hermite space $\mathcal{H}_{p_{\alpha, \gamma}}$.

According to Theorem 3.4 the reproducing kernel of $\mathcal{H}_{p\alpha,\gamma}$ is of the form

$$K_{p\alpha,\gamma}(x, y) = 1 + \gamma \sum_{k=1}^{\infty} k^{-\alpha} H_k(x) H_k(y).$$

The next theorem states sufficient conditions for a function to be in the Hermite space $\mathcal{H}_{p\alpha,\gamma}$ of functions on the \mathbb{R} . These conditions concern both the smoothness and the asymptotic behavior of the function at infinity. The theorem further gives a relation between the “smoothness parameter” α of the Hermite space and the actual smoothness of the function.

Theorem 3.5. *Let $\beta > 2$ be an integer and $f : \mathbb{R} \rightarrow \mathbb{R}$ be a β -times differentiable function satisfying*

- (i) $\mathcal{D}_x^j f(x) \varphi(x)^{\frac{1}{2}}$ is Lebesgue integrable for each $j \in \{1, \dots, \beta\}$ and
- (ii) $\mathcal{D}_x^j f(x) = O(e^{x^2/(2c)})$ as $|x| \rightarrow \infty$ for each $j \in \{0, \dots, \beta - 1\}$ and some $c > 1$,

where \mathcal{D}_x is the differential operator defined by (5). Then $f \in \mathcal{H}_{p\alpha,\gamma}$ for any $\alpha \in (1, \beta - 1)$ and any $\gamma > 0$.

Proof. Using integration by parts we obtain

$$\begin{aligned} \hat{f}(k) &= \int_{\mathbb{R}} f(x) H_k(x) \varphi(x) dx \\ &= \lim_{a \rightarrow \infty} f(x) \varphi(x) \frac{H_{k+1}(x)}{\sqrt{k+1}} \Big|_{x=-a}^a - \int_{\mathbb{R}} \frac{d}{dx} (f(x) \varphi(x)) \frac{H_{k+1}(x)}{\sqrt{k+1}} dx. \end{aligned}$$

Because of assumption (ii) there is a constant $\hat{c} > 0$ such that

$$|f(a) \varphi(a) H_{k+1}(a)| \leq \hat{c} e^{(\frac{1}{c}-1)\frac{a^2}{2}} |H_{k+1}(a)|$$

and, since $\frac{1}{c} - 1 < 0$ and $H_{k+1}(a)$ is a polynomial, we get that

$$\lim_{|a| \rightarrow \infty} f(a) \varphi(a) H_{k+1}(a) = 0.$$

Thus,

$$\begin{aligned} \hat{f}(k) &= -\frac{1}{\sqrt{k+1}} \int_{\mathbb{R}} \left(\frac{d}{dx} f(x) - x f(x) \right) H_{k+1}(x) \varphi(x) dx \\ &= -\frac{1}{\sqrt{k+1}} \int_{\mathbb{R}} \mathcal{D}_x f(x) H_{k+1}(x) \varphi(x) dx \\ &= -\frac{1}{\sqrt{k+1}} \widehat{\mathcal{D}_x f}(k+1). \end{aligned}$$

Now we can proceed in the same way, but for $\mathcal{D}_x^j f$, $j = 1, \dots, \beta - 1$, instead of f , such that after $\beta - 1$ times we end up with

$$\hat{f}(k) = (-1)^\beta \left(\frac{k!}{(k+\beta)!} \right)^{\frac{1}{2}} \widehat{\mathcal{D}_x^\beta f}(k+\beta).$$

So the k -th Hermite coefficient of f can be estimated by

$$\begin{aligned} |\hat{f}(k)| &\leq \left(\frac{k!}{(k+\beta)!} \right)^{\frac{1}{2}} \int_{\mathbb{R}} |\mathcal{D}^\beta f(x)| |H_{k+\beta}(x)| \varphi(x) dx \\ &\leq \left(\frac{k!}{(k+\beta)!} \right)^{\frac{1}{2}} \int_{\mathbb{R}} |\mathcal{D}^\beta f(x) \varphi(x)^{\frac{1}{2}}| dx \\ &= C \left(\frac{k!}{(k+\beta)!} \right)^{\frac{1}{2}} \end{aligned}$$

where $C = \int_{\mathbb{R}} |\mathcal{D}_x^\beta f(x) \varphi(x)^{\frac{1}{2}}| dx < \infty$ and Cramer's bound is applied for obtaining the second estimate. Then the norm of f can be bounded,

$$\begin{aligned} \|f\|_{p_{\alpha,\gamma}}^2 &= \sum_{k \in \mathbb{N}_0} p_{\alpha,\gamma}(k) |\hat{f}(k)|^2 \\ &\leq |\hat{f}(0)|^2 + \gamma^{-1} C^2 \sum_{k=1}^{\infty} k^\alpha \frac{k!}{(k+\beta)!} \\ &\leq |\hat{f}(0)|^2 + \gamma^{-1} C^2 \sum_{k=1}^{\infty} k^{-\beta+\alpha} \\ &= |\hat{f}(0)|^2 + \gamma^{-1} C^2 \zeta(\beta - \alpha) < \infty, \end{aligned}$$

if $\beta - \alpha > 1$. Therefore, we have that f is in the Hermite space $\mathcal{H}_{p_{\alpha,\gamma}}$ for any $\alpha \in (1, \beta - 1)$. \square

We can also show a reverse statement, i.e., the space $\mathcal{H}_{p_{\alpha,\gamma}}$ contains functions which are differentiable up to some finite order.

Theorem 3.6. *Let $f \in \mathcal{H}_{p_{\alpha,\gamma}}$. Then $f \in C^\beta(\mathbb{R}^d)$ for every $\beta < \alpha - 1$.*

Proof. For $n \in \mathbb{N}$ we write $f_n(x) = \sum_{k=0}^n \hat{f}(k) H_k(x)$ and let $\ell \in \mathbb{N}$ with $\ell \leq \beta$. We have that

$$\frac{d^\ell}{dx^\ell} f_n(x) = \sum_{k=0}^n \hat{f}(k) \frac{d^\ell}{dx^\ell} H_k(x) = \sum_{k=\ell}^n \hat{f}(k) \sqrt{\frac{k!}{(k-\ell)!}} H_{k-\ell}(x).$$

Then, for any $n \in \mathbb{N}$,

$$\begin{aligned} \left| \frac{d^\ell}{dx^\ell} f_n(x) \right| &= \left| \sum_{k=\ell}^n \left(\hat{f}(k) p_{\alpha,\gamma}(k)^{-\frac{1}{2}} \right) \left(p_{\alpha,\gamma}(k)^{\frac{1}{2}} \sqrt{\frac{k!}{(k-\ell)!}} H_{k-\ell}(x) \right) \right| \\ &\leq \|f\|_{p_{\alpha,\gamma}} \frac{1}{\varphi(x)} \left(\sum_{k=\ell}^n p_{\alpha,\gamma}(k) \frac{k!}{(k-\ell)!} \right)^{\frac{1}{2}} \\ &\leq \|f\|_{p_{\alpha,\gamma}} \frac{\gamma}{\varphi(x)} \left(\sum_{k=\ell}^n k^{-\alpha} \frac{k!}{(k-\ell)!} \right)^{\frac{1}{2}} \end{aligned}$$

Since $k!/(k-\ell)! \leq k^\ell$, we get

$$\left| \frac{d^\ell}{dx^\ell} f_n(x) \right| \leq \|f\|_{p_{\alpha,\gamma}} \frac{\gamma}{\varphi(x)} \left(\sum_{k=1}^n k^{-(\alpha-\ell)} \right)^{\frac{1}{2}}.$$

Moreover, we have that

$$\sum_{k=1}^{\infty} k^{-(\alpha-\ell)} = \zeta(\alpha-\ell) < \infty,$$

because $\alpha-\ell \geq \alpha-\beta > \alpha-(\alpha-1) > 1$. Thus, we know that the sequence $(\frac{d^\ell}{dx^\ell} f_n(x))_{n \in \mathbb{N}}$ converges uniformly on compacts. Furthermore, from Theorem 3.2 it follows that f_n converges pointwise to f . Because of a basic theorem about uniform convergence and differentiation, see e.g., Rudin [16, Theorem 7.17] we get that $\frac{d^\ell}{dx^\ell} f$ exists and it is given by

$$\frac{d^\ell}{dx^\ell} f(x) = \sum_{k=0}^{\infty} \widehat{f}(k) \sqrt{\frac{k!}{(k-\ell)!}} H_{k-\ell}(x).$$

The continuity of the derivatives also follows by a basic theorem about uniform convergence, see e.g., Rudin [16, Theorem 7.12]. \square

Now, we suggest to generalize this type of Hermite space to functions on the \mathbb{R}^d in the following manner. Let $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)$ with $\gamma_j > 0$ be non-increasing weights and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ with $\alpha_j > 1$ smoothness parameters. Then we define the function $\mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}} : \mathbb{N}_0^d \longrightarrow \mathbb{R}^+$ by

$$\mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\mathbf{k}) := \prod_{j=1}^d p_{\alpha_j, \gamma_j}(k_j), \quad (8)$$

which is summable, because

$$\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\mathbf{k}) = \prod_{j=1}^d (1 + \gamma_j \zeta(\alpha_j)) < \infty. \quad (9)$$

So we get the Hermite space $\mathcal{H}_{\mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}}$ which also can be written as the d -fold Hilbert space tensor product of Hermite spaces of functions on the \mathbb{R} ,

$$\mathcal{H}_{\mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}} = \mathcal{H}_{p_{\alpha_1, \gamma_1}} \otimes \dots \otimes \mathcal{H}_{p_{\alpha_d, \gamma_d}}.$$

The reproducing kernel $K_{\mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}}$ is the product of the kernels $K_{p_{\alpha_i, \gamma_i}}$,

$$\begin{aligned} K_{\mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}}(\mathbf{x}, \mathbf{y}) &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{p}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) H_{\mathbf{k}}(\mathbf{y}) \\ &= \prod_{j=1}^d \left(1 + \gamma_j \sum_{k=1}^{\infty} k^{-\alpha_j} H_k(x_j) H_k(y_j) \right) \end{aligned}$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. The Hermite coefficients of the functions of $\mathcal{H}_{\mathbf{p}_{\alpha}, \gamma}$ again have to decay polynomially and with the smoothness parameters α one can control the order of the polynomial decay for each direction separately. The weights γ determine the influence of the coordinates. Since we assumed that γ is a non-increasing sequence, the weights moderate the influence of the coordinates. That means that each coordinate is less or at most equally important than the previous coordinates. Because of the product-form of the function $\mathbf{p}_{\alpha, \gamma}$ we are dealing with so-called product weights, first introduced in Sloan and Woźniakowski [18]. Because of the weights γ we also call $\mathcal{H}_{\mathbf{p}_{\alpha}, \gamma}$ a *weighted Hermite space*. Furthermore, note that Theorem 3.5 can be extended to Hermite spaces $\mathcal{H}_{\mathbf{p}_{\alpha}, \gamma}$ of functions on the \mathbb{R}^d which means that the smoothness-parameters α correspond to the differentiability of the functions.

3.1.2 Weighted Hermite spaces with exponentially decaying coefficients

Let $\gamma = (\gamma_1, \dots, \gamma_d)$ be a sequence of non-increasing positive weights and $\omega = (\omega_1, \dots, \omega_d) \in (0, 1)^d$. We define a function $\epsilon_{\omega, \gamma} : \mathbb{N}_0^d \longrightarrow \mathbb{R}^+$ as

$$\epsilon_{\omega, \gamma}(\mathbf{k}) = \prod_{j=1}^d \epsilon_{\omega_j, \gamma_j}(k_j) \quad (10)$$

with

$$\epsilon_{\omega_j, \gamma_j}(k_j) = \begin{cases} 1 & \text{if } k_j = 0, \\ \gamma_j \omega_j^{k_j} & \text{if } k_j \neq 0 \end{cases}$$

for $j = 1, \dots, d$. Since

$$\begin{aligned} \sum_{\mathbf{k} \in \mathbb{N}_0^d} \epsilon_{\omega, \gamma}(\mathbf{k}) &= \prod_{j=1}^d \left(1 + \gamma_j \sum_{k=1}^{\infty} \omega_j^k \right) \\ &= \prod_{j=1}^d \left(1 + \gamma_j \frac{\omega_j}{1 - \omega_j} \right) < \infty, \end{aligned}$$

we have that $\epsilon_{\omega, \gamma}$ is summable and we get the Hermite space $\mathcal{H}_{\epsilon_{\omega, \gamma}}$ which has the Hilbert space tensor product form

$$\mathcal{H}_{\epsilon_{\omega, \gamma}} = \mathcal{H}_{\epsilon_{\omega_1, \gamma_1}} \otimes \dots \otimes \mathcal{H}_{\epsilon_{\omega_d, \gamma_d}}.$$

The Hermite coefficients of the elements of $\mathcal{H}_{\epsilon_{\omega, \gamma}}$ decay exponentially which is controlled by the parameter ω . We again have product weights and the influence of the coordinates can be determined by the choice of the weights $(\gamma_1, \dots, \gamma_d)$.

The reproducing kernel of $\mathcal{H}_{\epsilon_{\omega, \gamma}}$ is given by

$$K_{\epsilon_{\omega, \gamma}}(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^d \left(1 + \gamma_j \sum_{k=1}^{\infty} \omega_j^k H_k(x_j) H_k(y_j) \right),$$

which follows from Theorem 3.4. From Mehler's formula, see Szegő [19], we can further derive, for each $j = 1, \dots, d$, the following formula

$$1 + \gamma_j \sum_{k=1}^{\infty} \omega_j^k H_k(x_j) H_k(y_j) = 1 - \gamma_j + \gamma_j \frac{1}{\sqrt{1 - \omega_j^2}} \exp \left(\frac{\omega_j}{1 + \omega_j} x_j y_j - \frac{\omega_j^2}{2(1 - \omega_j^2)} (x_j - y_j)^2 \right)$$

and thus,

$$K_{\epsilon_{\omega}, \gamma}(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^d \left(1 - \gamma_j + \gamma_j \frac{1}{\sqrt{1 - \omega_j^2}} \exp \left(\frac{\omega_j}{1 + \omega_j} x_j y_j - \frac{\omega_j^2}{2(1 - \omega_j^2)} (x_j - y_j)^2 \right) \right). \quad (11)$$

From this representation of the reproducing kernel we see that for $\gamma_j < 1$, $j = 1, \dots, d$, the reproducing kernel $K_{\epsilon_{\omega}, \gamma}$ is positive.

We want to know more about the functions that are contained in the Hermite space $\mathcal{H}_{\epsilon_{\omega}, \gamma}$. In the previous subsection, Theorem 3.5 gave us sufficient conditions on a function such that the function belongs to the Hermite space of functions with polynomially decaying Hermite coefficients. Unfortunately, we do not have a similar statement concerning the Hermite space of exponential decaying Hermite coefficients. Nevertheless, we know that the Hermite space $\mathcal{H}_{\epsilon_{\omega}, \gamma}$ contains all polynomials, because $\text{span}\{H_{\mathbf{k}} : |\mathbf{k}| \leq m\}$ is equal to the set of all polynomials on the \mathbb{R}^d with degree smaller or equal to m . Furthermore, the function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, given by $f(\mathbf{x}) = f(x_1, \dots, x_d) = \exp(\mathbf{w}^\top \mathbf{x})$ with $\mathbf{w} \in \mathbb{R}^d$, belongs to $\mathcal{H}_{\epsilon_{\omega}, \gamma}$ for any weight sequence γ and any ω . This is true, because we have that the \mathbf{k} -th Hermite coefficient is given by

$$\hat{f}(\mathbf{k}) = e^{\mathbf{w}^\top \mathbf{w}} \frac{\mathbf{w}^{\mathbf{k}}}{\sqrt{\mathbf{k}!}}$$

and consequently, the norm of f is finite,

$$\|f\|_{\epsilon_{\omega}, \gamma}^2 = e^{\mathbf{w}^\top \mathbf{w}} \prod_{j=1}^d \left(1 + \gamma_j \sum_{k=1}^{\infty} \omega_j^k \frac{w_j^{2k}}{k!} \right) \leq e^{\mathbf{w}^\top \mathbf{w}} \prod_{j=1}^d \exp(\omega_j w_j^2) < \infty.$$

In addition, we can show that the functions in $\mathcal{H}_{\epsilon_{\omega}, \gamma}$ are analytic, i.e., analyticity is a necessary condition on the functions for being in the Hermite space of exponentially decaying Hermite coefficients.

Proposition 3.7. *Let $f \in \mathcal{H}_{\epsilon_{\omega}, \gamma}$. Then f is analytic.*

Proof. Since $\gamma_j \leq 1$ for all $j = 1, \dots, d$, we have $\mathcal{H}_{\epsilon_{\omega}, \gamma} \subseteq \mathcal{H}_{\epsilon_{\omega}, \mathbf{1}}$ with $\mathbf{1} = (1, \dots, 1)$. Therefore, it is sufficient to show analyticity for functions f which belong to $\mathcal{H}_{\epsilon_{\omega}, \mathbf{1}}$. Let $\ell \in \mathbb{N}_0^d$ and $f \in \mathcal{H}_{\epsilon_{\omega}, \mathbf{1}}$. Analogue to the proof of Proposition 3.6 it can be shown that the derivative $\frac{\partial^{|\ell|}}{\partial \mathbf{x}^\ell} f$ exists and that we can interchange summation and differentiation, i.e., we obtain

$$\frac{\partial^{|\ell|}}{\partial \mathbf{x}^\ell} f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) \frac{\partial^{|\ell|}}{\partial \mathbf{x}^\ell} H_{\mathbf{k}}(\mathbf{x}) = \sum_{\mathbf{k} \geq \ell} \hat{f}(\mathbf{k}) \sqrt{\frac{\mathbf{k}!}{(\mathbf{k} - \ell)!}} H_{\mathbf{k} - \ell}(\mathbf{x}).$$

Then,

$$\begin{aligned}
\left| \frac{\partial^{|\ell|}}{\partial \mathbf{x}^\ell} f(\mathbf{x}) \right| &= \left| \sum_{\mathbf{k} \geq \ell} \left(\widehat{f}(\mathbf{k}) (\epsilon_{\omega,1}(\mathbf{k}))^{-\frac{1}{2}} \right) \left((\epsilon_{\omega,1}(\mathbf{k}))^{\frac{1}{2}} \sqrt{\frac{\mathbf{k}!}{(\mathbf{k}-\ell)!}} H_{\mathbf{k}-\ell}(\mathbf{x}) \right) \right| \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \left(\sum_{\mathbf{k} \geq \ell} \epsilon_{\omega,1}(\mathbf{k}) \frac{\mathbf{k}!}{(\mathbf{k}-\ell)!} \right)^{\frac{1}{2}} \\
&= \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \left(\sum_{\mathbf{k} \geq \ell} \prod_{j=1}^d (\ell_j!) \omega_j^{k_j} \frac{k_j!}{(k_j - \ell_j)! (\ell_j!)} \right)^{\frac{1}{2}} \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \prod_{j=1}^d \sqrt{\ell_j!} \left(\sum_{k=\ell_j}^{\infty} \binom{k}{\ell_j} \omega_j^k \right)^{\frac{1}{2}} \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \sqrt{\ell!} \prod_{j=1}^d \left(\frac{\omega_j^{\ell_j}}{(1 - \omega_j)^{\ell_j+1}} \right)^{\frac{1}{2}} \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \sqrt{\ell!} \prod_{j=1}^d \left(\frac{(\omega^*)^{\ell_j}}{(1 - \omega^*)^{\ell_j+1}} \right)^{\frac{1}{2}}
\end{aligned}$$

where $\omega^* = \max_j \omega_j$. Now we show that f can locally be represented by its Taylor expansion. For any $\mathbf{y} \in \mathbb{R}^d$ and any $\mathbf{x} \in \mathbb{R}^d$ with $\|\mathbf{x} - \mathbf{y}\|_\infty < \frac{1-\omega^*}{\omega^*}$,

$$\begin{aligned}
\left| \sum_{\ell \in \mathbb{N}_0^d} \frac{1}{\ell!} \frac{\partial^{|\ell|}}{\partial \mathbf{x}^\ell} f(\mathbf{y}) \prod_{j=1}^d (x_j - y_j)^{\ell_j} \right| &\leq \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \sum_{\ell \in \mathbb{N}_0^d} \prod_{j=1}^d \frac{1}{\sqrt{\ell_j!}} \left(\frac{(\omega^*)^{\ell_j} (x_j - y_j)^{2\ell_j}}{(1 - \omega^*)^{\ell_j+1}} \right)^{\frac{1}{2}} \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \prod_{j=1}^d \sum_{\ell=0}^{\infty} \frac{1}{\sqrt{\ell!}} \left(\frac{(\omega^*)^\ell (x_j - y_j)^{2\ell}}{(1 - \omega^*)^{\ell+1}} \right)^{\frac{1}{2}} \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \prod_{j=1}^d \left(\sum_{\ell=0}^{\infty} \frac{1}{\ell!} \right)^{\frac{1}{2}} \left(\sum_{\ell=0}^{\infty} \frac{(\omega^*)^\ell (x_j - y_j)^{2\ell}}{(1 - \omega^*)^{\ell+1}} \right)^{\frac{1}{2}} \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \left(\frac{e}{1 - \omega^*} \sum_{\ell=0}^{\infty} \left(\frac{\omega^* \|\mathbf{x} - \mathbf{y}\|_\infty^2}{1 - \omega^*} \right)^\ell \right)^{d/2} \\
&\leq \|f\|_{\epsilon_{\omega,1}} \frac{1}{\varphi_d(\mathbf{x})} \left(\frac{e}{1 - \omega^* - \omega^* \|\mathbf{x} - \mathbf{y}\|_\infty^2} \right)^d < \infty.
\end{aligned}$$

Thus, we have that the Taylor expansion converges locally. It remains to show that the remainder R_n of the Taylor polynomial, given by

$$R_n := \sum_{|\mathbf{k}|=n+1} \frac{n+1}{\mathbf{k}!} (\mathbf{x} - \mathbf{y})^{\mathbf{k}} \int_0^1 (1-t)^n \frac{\partial^{|\mathbf{k}|}}{\partial \mathbf{x}^{\mathbf{k}}} f(\mathbf{y} + t(\mathbf{x} - \mathbf{y})) dt,$$

vanishes if n goes to infinity. We have

$$\begin{aligned}
|R_n| &\leq \sum_{|\mathbf{k}|=n+1} \frac{n+1}{\mathbf{k}!} |\mathbf{x} - \mathbf{y}|^{\mathbf{k}} \int_0^1 |1-t|^n \left| \frac{\partial^{|\mathbf{k}|}}{\partial \mathbf{x}^{\mathbf{k}}} f(\mathbf{y} + t(\mathbf{x} - \mathbf{y})) \right| dt \\
&\leq \sum_{|\mathbf{k}|=n+1} (n+1) |\mathbf{x} - \mathbf{y}|^{\mathbf{k}} \int_0^1 \frac{\|f\|_{\epsilon_{\omega,1}} |1-t|^n}{\varphi_d(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))} \prod_{j=1}^d \left(\frac{(\omega^*)^{k_j}}{k_j! (1 - \omega^*)^{k_j+1}} \right)^{\frac{1}{2}} dt \\
&\leq (n+1) \left[\int_0^1 \frac{\|f\|_{\epsilon_{\omega,1}} |1-t|^n}{\varphi_d(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))} dt \right] \sum_{|\mathbf{k}|=n+1} \prod_{j=1}^d \left(\frac{(\omega^*)^{k_j} |x_j - y_j|^{2k_j}}{(1 - \omega^*)^{k_j+1}} \right)^{\frac{1}{2}}.
\end{aligned}$$

Since $\|\mathbf{x} - \mathbf{y}\|_{\infty} < \sqrt{\frac{1-\omega^*}{\omega^*}}$, we have for any $j = 1, \dots, d$,

$$\frac{1}{\varphi(y_j + t(x_j - y_j))} \leq \begin{cases} 1/\varphi(y_j + \sqrt{(1-\omega^*)/\omega^*}) & \text{if } y_j \geq 0; \\ 1/\varphi(y_j - \sqrt{(1-\omega^*)/\omega^*}) & \text{if } y_j < 0, \end{cases}$$

such that we can bound $1/\varphi_d(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))$ by some constant C_1 independent of n and t . Hence,

$$\begin{aligned}
|R_n| &\leq C_1 \|f\|_{\epsilon_{\omega,1}} (n+1) \left[\int_0^1 |1-t|^n dt \right] \sum_{|\mathbf{k}|=n+1} \prod_{j=1}^d \left(\frac{(\omega^*)^{k_j} |x_j - y_j|^{2k_j}}{(1 - \omega^*)^{k_j+1}} \right)^{\frac{1}{2}} \\
&\leq C_1 \|f\|_{\epsilon_{\omega,1}} \frac{1}{(1 - \omega^*)^{d/2}} \sum_{|\mathbf{k}|=n+1} \prod_{j=1}^d \left(\frac{\omega^* \|\mathbf{x} - \mathbf{y}\|_{\infty}^2}{1 - \omega^*} \right)^{\frac{k_j}{2}} \\
&\leq C_1 \|f\|_{\epsilon_{\omega,1}} \frac{1}{(1 - \omega^*)^{d/2}} \sum_{|\mathbf{k}|=n+1} \left(\frac{\omega^* \|\mathbf{x} - \mathbf{y}\|_{\infty}^2}{1 - \omega^*} \right)^{\frac{|\mathbf{k}|}{2}} \\
&\leq \frac{C_1 \|f\|_{\epsilon_{\omega,1}}}{(1 - \omega^*)^{d/2}} \left(\frac{\omega^* \|\mathbf{x} - \mathbf{y}\|_{\infty}^2}{1 - \omega^*} \right)^{\frac{n+1}{2}} \frac{(d+n)!}{(d-1)!(n+1)!}.
\end{aligned}$$

Since $\omega^* \|\mathbf{x} - \mathbf{y}\|_{\infty}^2 / (1 - \omega^*) < 1$ and $(d+n)! / ((d-1)!(n+1)!) = \mathcal{O}(n^{d-1})$, we get that $R_n \rightarrow 0$ as n goes to ∞ . Thus, f is indeed analytic. \square

3.2 Multivariate integration and tractability in weighted Hermite spaces

A survey of the analysis of multivariate integration in reproducing kernel Hilbert spaces is given, e.g., in Dick and Pillichshammer [7], where they study the integration of functions over the unit cube $[0, 1]^d$. In the following, we intend to study multivariate integration of functions which belong to a Hermite space $\mathcal{H}_{\mathbf{r}}$. For that, we consider the integral of functions $f \in \mathcal{H}_{\mathbf{r}}$ over the \mathbb{R}^d with respect to the Gaussian measure, i.e.,

$$I(f) = \int_{\mathbb{R}^d} f(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x}. \tag{12}$$

The linear functional I is well-defined, because a Gaussian square-integrable function f also is integrable with respect to the Gaussian measure. Now we intend to approximate the d -

dimensional integral (12) by a quasi-Monte Carlo rule, i.e., an equally weighted quadrature rule

$$Q_{n,d}(f, \mathcal{P}) = \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i)$$

with deterministic point set $\mathcal{P} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^d$. For that, it is interesting to analyze the integration error of a QMC rule which depends on both the function f and the point set \mathcal{P} .

Definition 3.8. Let \mathcal{H}_r be a Hermite space and let \mathcal{P} be the quadrature points used in the QMC rule $Q_{n,d}$.

(i) For $f \in \mathcal{H}_r$ the *QMC integration error* of f is given by

$$e(f, \mathcal{P}) := |I(f) - Q_{n,d}(f, \mathcal{P})|.$$

(ii) The *worst case error* for QMC integration in \mathcal{H}_r is defined as

$$e_{n,d}(\mathcal{P}, \mathcal{H}_r) := \sup_{f \in \mathcal{H}_r, \|f\|_r \leq 1} e(f, \mathcal{P}) = \sup_{f \in \mathcal{H}_r, \|f\|_r \leq 1} |I(f) - Q_{n,d}(f, \mathcal{P})|.$$

Since we know that a Hermite space has a reproducing kernel function K_r , we can use the reproducing property of K_r to get

$$\int_{\mathbb{R}^d} f(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} \langle f, K_r(\cdot, \mathbf{x}) \rangle_r \varphi_d(\mathbf{x}) d\mathbf{x} = \left\langle f, \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \right\rangle_r,$$

where we have used continuity of integration on the space \mathcal{H}_r , which in turn follows from

$$\left| \int_{\mathbb{R}^d} f(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \right| = |\hat{f}(0)| \leq \|f\|_r.$$

Therefore the *representer of integration* is the function $\mathbf{x} \mapsto \int_{\mathbb{R}^d} K_r(\mathbf{x}, \mathbf{y}) \varphi_d(\mathbf{y}) d\mathbf{y}$. In the same manner we obtain

$$\frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i) = \frac{1}{n} \sum_{i=1}^n \langle f, K_r(\cdot, \mathbf{x}_i) \rangle_r = \left\langle f, \frac{1}{n} \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i) \right\rangle_r,$$

that means, $\mathbf{x} \mapsto \frac{1}{n} \sum_{i=1}^n K_r(\mathbf{x}, \mathbf{x}_i)$ is the representer of the QMC rule. Now, the integration error $e(f, \mathcal{P})$ can be estimated by

$$\begin{aligned} |I(f) - Q_{n,d}(f, \mathcal{P})| &= \left| \left\langle f, \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i) \right\rangle_r \right| \\ &\leq \|f\|_r \left\| \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i) \right\|_r. \end{aligned}$$

So for any point set $\mathcal{P} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ the worst case error $e_{n,d}(\mathcal{P}, \mathcal{H}_r)$ for QMC integration in

the Hermite space \mathcal{H}_r is given by

$$e_{n,d}(\mathcal{P}, \mathcal{H}_r) = \left\| \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i) \right\|_r$$

Because of the linearity of the inner product we obtain

$$\begin{aligned} e_{n,d}^2(\mathcal{P}, \mathcal{H}_r) &= \left\langle \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i), \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i) \right\rangle_r \\ &= \left\langle \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x}, \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \right\rangle_r - \frac{2}{n} \left\langle \int_{\mathbb{R}^d} K_r(\cdot, \mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x}, \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i) \right\rangle_r \\ &\quad + \frac{1}{n^2} \left\langle \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i), \sum_{i=1}^n K_r(\cdot, \mathbf{x}_i) \right\rangle_r \end{aligned}$$

and further

$$\begin{aligned} e_{n,d}^2(\mathcal{P}, \mathcal{H}_r) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left\langle K_r(\cdot, \mathbf{x}), K_r(\cdot, \mathbf{y}) \right\rangle_r \varphi_d(\mathbf{x}) \varphi_d(\mathbf{y}) d\mathbf{x} d\mathbf{y} - \frac{2}{n} \sum_{i=1}^n \int_{\mathbb{R}^d} \left\langle K_r(\cdot, \mathbf{x}), K_r(\cdot, \mathbf{x}_i) \right\rangle_r \varphi_d(\mathbf{x}) d\mathbf{x} \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left\langle K_r(\cdot, \mathbf{x}_i), K_r(\cdot, \mathbf{x}_j) \right\rangle_r. \end{aligned}$$

Using the reproducing property of K_r we end up with

$$\begin{aligned} e_{n,d}^2(\mathcal{P}, \mathcal{H}_r) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_r(\mathbf{x}, \mathbf{y}) \varphi_d(\mathbf{x}) \varphi_d(\mathbf{y}) d\mathbf{x} d\mathbf{y} - \frac{2}{n} \sum_{i=1}^n \int_{\mathbb{R}^d} K_r(\mathbf{x}, \mathbf{x}_i) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_r(\mathbf{x}_i, \mathbf{x}_j). \end{aligned}$$

If we now use that $K_r(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) H_{\mathbf{k}}(\mathbf{y})$, we get for the first term

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_r(\mathbf{x}, \mathbf{y}) \varphi_d(\mathbf{x}) \varphi_d(\mathbf{y}) d\mathbf{x} d\mathbf{y} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) H_{\mathbf{k}}(\mathbf{y}) \varphi_d(\mathbf{x}) \varphi_d(\mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{y}) \varphi_d(\mathbf{y}) d\mathbf{y} \\ &= \mathbf{r}(\mathbf{0}) \end{aligned}$$

and for the second term

$$\begin{aligned} \frac{2}{n} \sum_{i=1}^n \int_{\mathbb{R}^d} K_r(\mathbf{x}, \mathbf{x}_i) \varphi_d(\mathbf{x}) d\mathbf{x} &= \frac{2}{n} \sum_{i=1}^n \int_{\mathbb{R}^d} \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}) H_{\mathbf{k}}(\mathbf{x}_i) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= \frac{2}{n} \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) \sum_{i=1}^n H_{\mathbf{k}}(\mathbf{x}_i) \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \end{aligned}$$

$$= \frac{2}{n} \mathbf{r}(\mathbf{0}) \sum_{i=1}^n H_0(\mathbf{x}_i) = 2\mathbf{r}(\mathbf{0}).$$

Therefore, we obtain

$$e_{n,d}(\mathcal{P}, \mathcal{H}_r) = \left(-\mathbf{r}(\mathbf{0}) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_r(\mathbf{x}_i, \mathbf{x}_j) \right)^{\frac{1}{2}}. \quad (13)$$

To get an estimate for the integration error, we would like to have an upper bound for the worst case error $e_{n,d}(\mathcal{P}, \mathcal{H}_r)$. Since we have not chosen a specific point set \mathcal{P} , we derive an upper bound through an averaging argument.

Definition 3.9. The Gaussian weighted root-mean-square error for QMC integration is

$$\bar{e}_{n,d}(\mathcal{H}_r) := \left(\int_{\mathbb{R}^{dn}} e_{n,d}^2(\mathbf{x}_1, \dots, \mathbf{x}_n; \mathcal{H}_r) \varphi_d(\mathbf{x}_1) \dots \varphi_d(\mathbf{x}_n) d(\mathbf{x}_1, \dots, \mathbf{x}_n) \right)^{\frac{1}{2}}$$

where $e_{n,d}(\mathbf{x}_1, \dots, \mathbf{x}_n; \mathcal{H}_r)$ is the worst case error of the point set $\mathcal{P} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$.

Proposition 3.10. Let \mathcal{H}_r be a Hermite space. Then the Gaussian weighted root-mean-square error $\bar{e}_{n,d}(\mathcal{H}_r)$ for QMC integration is given by

$$\bar{e}_{n,d}(\mathcal{H}_r) = \frac{1}{\sqrt{n}} \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) - \mathbf{r}(\mathbf{0}) \right)^{\frac{1}{2}}.$$

Proof. With (13) the Gaussian weighted root-mean-square error can be written as

$$\bar{e}_{n,d}^2(\mathcal{H}_r) = -\mathbf{r}(\mathbf{0}) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_r(\mathbf{x}_i, \mathbf{x}_j) \varphi_d(\mathbf{x}_i) \varphi_d(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j$$

Now we consider

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_r(\mathbf{x}_i, \mathbf{x}_j) \varphi_d(\mathbf{x}_i) \varphi_d(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j &= \\ &= \sum_{i=1}^n \int_{\mathbb{R}^d} K_r(\mathbf{x}_i, \mathbf{x}_i) \varphi_d(\mathbf{x}_i) d\mathbf{x}_i + \sum_{i \neq j} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_r(\mathbf{x}_i, \mathbf{x}_j) \varphi_d(\mathbf{x}_i) \varphi_d(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j \end{aligned}$$

in more detail. Since

$$\begin{aligned} \int_{\mathbb{R}^d} K_r(\mathbf{x}_i, \mathbf{x}_i) \varphi_d(\mathbf{x}_i) d\mathbf{x}_i &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x})^2 \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) \end{aligned}$$

and for $i \neq j$

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_{\mathbf{r}}(\mathbf{x}_i, \mathbf{x}_j) \varphi_d(\mathbf{x}_i) \varphi_d(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j &= \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \int_{\mathbb{R}^d} H_{\mathbf{k}}(\mathbf{y}) \varphi_d(\mathbf{y}) d\mathbf{y} \\ &= \mathbf{r}(\mathbf{0}), \end{aligned}$$

we have that

$$\sum_{i=1}^n \sum_{j=1}^n \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_{\mathbf{r}}(\mathbf{x}_i, \mathbf{x}_j) \varphi_d(\mathbf{x}_i) \varphi_d(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j = n \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) + (n^2 - n) \mathbf{r}(\mathbf{0}).$$

So we end up with

$$\begin{aligned} \bar{e}_{n,d}^2(\mathcal{H}_{\mathbf{r}}) &= -\mathbf{r}(\mathbf{0}) + \frac{1}{n} \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) + \frac{n^2 - n}{n^2} \mathbf{r}(\mathbf{0}) \\ &= \frac{1}{n} \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\mathbf{k}) - \mathbf{r}(\mathbf{0}) \right). \end{aligned}$$

□

Corollary 3.11. *Let $\mathcal{H}_{\mathbf{r}}$ be Hermite space of d -dimensional functions. For each $n \in \mathbb{N}$ there exists a point set $\mathcal{P} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ such that for the worst case error for QMC integration in $\mathcal{H}_{\mathbf{r}}$ the upper bound*

$$e_{d,n}(\mathcal{P}, \mathcal{H}_{\mathbf{r}}) \leq \frac{c(d)}{\sqrt{n}}$$

holds, where $c(d) := \sqrt{\sum_{\mathbf{k} \in \mathbb{N}_0^d \setminus \{\mathbf{0}\}} \mathbf{r}(\mathbf{k})}$.

Proof. We can estimate the minimum of the worst case error over all possible point sets \mathcal{P} by the Gaussian weighted root-mean-square error, i.e.,

$$\min_{\mathcal{P}=\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^d} e_{n,d}(\mathcal{P}, \mathcal{H}_{\mathbf{r}}) \leq \bar{e}_{n,d}(\mathcal{H}_{\mathbf{r}}).$$

With Proposition 3.10 we get

$$\min_{\mathcal{P}=\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^d} e_{n,d}(\mathcal{P}) \leq \frac{c(d)}{\sqrt{n}}.$$

□

Corollary 3.11 states that a sequence of point sets $(\mathcal{P}_n)_{n \in \mathbb{N}}$ can be chosen such that the convergence rate of the corresponding quasi-Monte Carlo method is $\frac{1}{2}$, which is the convergence rate of crude Monte Carlo simulation. However, this upper bound results from an averaging argument and, consequently, we are confident that the upper bound can be improved for specific point sets. Furthermore, $c(d)$ is constant with respect to n but it still can depend on the

dimension d and it depends on the function \mathbf{r} . To get more information about the behavior of $c(d)$, we need additional information about the function \mathbf{r} . For that, we will consider the two examples of Hermite spaces, given in the previous subsection.

To study the dependence on dimension d it is reasonable to consider the minimal number of information operations which are required to reduce the initial error, given by

$$e_{0,d}(\mathcal{H}_{\mathbf{r}}) = \sup_{f \in \mathcal{H}_{\mathbf{r}}, \|f\|_{\mathbf{r}} \leq 1} |I(f)|,$$

by a factor $\varepsilon \in (0, 1)$. Since $I(f) = \hat{f}(\mathbf{0}) = \langle f, 1 \rangle_{\mathbf{r}}$, we know that the initial error is $\mathbf{r}(\mathbf{0})$. Now let

$$n_{\min}(\varepsilon, d; \mathcal{H}_{\mathbf{r}}) := \min\{n : \exists \mathcal{P} \text{ such that } e_{n,d}(\mathcal{P}, \mathcal{H}_{\mathbf{r}}) \leq \varepsilon \mathbf{r}(\mathbf{0})\},$$

which is called the *information complexity* of quasi-Monte Carlo integration in the Hermite space $\mathcal{H}_{\mathbf{r}}$.

Definition 3.12. 1. Multivariate integration in the Hermite space $\mathcal{H}_{\mathbf{r}}$ is called *polynomially tractable* if there exist $c, p, q \in \mathbb{R}^+$ such that

$$n_{\min}(\varepsilon, d; \mathcal{H}_{\mathbf{r}}) \leq c d^q \varepsilon^{-p} \quad (14)$$

holds for all dimensions $d \in \mathbb{N}$ and for all $\varepsilon \in (0, 1)$. The infima of the numbers p and q are called ε - and d -exponents of polynomial tractability.

2. If (14) holds with $q = 0$, we say that multivariate integration in $\mathcal{H}_{\mathbf{r}}$ is strongly polynomially tractable. The infimum of p is called ε -exponent of strong polynomial tractability
3. If (14) does not hold, we say that multivariate integration in $\mathcal{H}_{\mathbf{r}}$ is polynomially intractable.

We start with the Hermite space $\mathcal{H}_{\mathbf{p}_{\alpha}, \gamma}$ of functions with polynomially decaying Hermite coefficients and we will show that the choice of the weights γ is crucial for obtaining polynomial tractability.

Proposition 3.13. *Let $\gamma = (\gamma_1, \dots, \gamma_d) \in (0, \infty)^d$ be a non-increasing sequence of weights and $\alpha = (\alpha_1, \dots, \alpha_d) \in (1, \infty)^d$ a sequence of smoothness parameters. There exists a quasi-Monte Carlo method with point set \mathcal{P} such that the worst case error for QMC integration in the Hermite space $\mathcal{H}_{\mathbf{p}_{\alpha}, \gamma}$ of functions with polynomially decaying coefficients can be bounded from above by*

$$e_{n,d}(\mathcal{P}, \mathcal{H}_{\mathbf{p}_{\alpha}, \gamma}) \leq \frac{1}{\sqrt{n}} e^{\frac{\zeta(\alpha_{\min})}{2} \sum_{j=1}^d \gamma_j} \quad (15)$$

with $\alpha_{\min} := \min\{\alpha_i : i = 1, \dots, d\}$.

Proof. From Corollary 3.11 we know that there exists a quasi-Monte Carlo method such that

the worst case error can be estimated by

$$e_{n,d}(\mathcal{P}, \mathcal{H}_{\mathbf{p}_{\alpha,\gamma}}) \leq \frac{1}{\sqrt{n}} \left(\sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{p}_{\alpha,\gamma}(\mathbf{k}) - 1 \right)^{\frac{1}{2}}.$$

Due to (9) we get

$$\begin{aligned} e_{n,d}(\mathcal{P}, \mathcal{H}_{\mathbf{p}_{\alpha,\gamma}}) &\leq \frac{1}{\sqrt{n}} \left(\prod_{j=1}^d (1 + \gamma_j \zeta(\alpha_j)) \right)^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{n}} e^{\frac{1}{2} \sum_{j=1}^d \ln(1 + \gamma_j \zeta(\alpha_j))} \\ &\leq \frac{1}{\sqrt{n}} e^{\frac{1}{2} \sum_{j=1}^d \gamma_j \zeta(\alpha_j)}. \end{aligned}$$

Denote the smallest smoothness parameter as α_{\min} . Then the statement follows,

$$e_{n,d}(\mathcal{P}, \mathcal{H}_{\mathbf{p}_{\alpha,\gamma}}) \leq \frac{1}{\sqrt{n}} e^{\frac{\zeta(\alpha_{\min})}{2} \sum_{j=1}^d \gamma_j}.$$

□

The upper bound (15) allows us to give sufficient conditions on the weights concerning the tractability of multivariate integration in the Hermite space $\mathcal{H}_{\mathbf{p}_{\alpha,\gamma}}$. Let γ be an infinite non-increasing sequence of positive weights and α be an infinite sequence of smoothness parameters. Furthermore, assume that there exist a $\alpha_{\min} > 1$ such that $\alpha_j > \alpha_{\min}$ holds for all $j \in \mathbb{N}$. From (15) we obtain that

$$n_{\min}(\varepsilon, d, \mathcal{H}_{\mathbf{p}_{\alpha,\gamma}}) \leq \varepsilon^{-2} e^{\zeta(\alpha_{\min}) \sum_{j=1}^d \gamma_j}.$$

In the case that $\sum_{j=1}^{\infty} \gamma_j < \infty$ holds, the upper bound of the information complexity does not depend on the dimension d and thus we get a sufficient condition for strong polynomial tractability. If we suppose that $\limsup_d \sum_{j=1}^d \gamma_j / \ln(d) < \infty$,

$$n_{\min}(\varepsilon, d; \mathcal{H}_{\mathbf{p}_{\alpha,\gamma}}) \leq \varepsilon^{-2} d^{\zeta(\alpha_{\min}) \sum_{j=1}^d \gamma_j / \ln(d)}.$$

Theorem 3.14. *Let $\gamma = (\gamma_1, \gamma_2, \dots) \in (0, \infty)^{\mathbb{N}}$ be a non-increasing sequence of weights and $\alpha = (\alpha_1, \alpha_2, \dots) \in (1, \infty)^{\mathbb{N}}$ be a sequence of smoothness parameters such that $\inf_d \alpha_d > 1$. Multivariate integration in the weighted Hermite space $\mathcal{H}_{\mathbf{p}_{\alpha,\gamma}}$ of functions with polynomially decaying coefficients is*

1. *strongly polynomially tractable, if $\sum_{j=1}^{\infty} \gamma_j < \infty$.*
2. *polynomially tractable, if $A := \limsup_d \frac{\sum_{j=1}^d \gamma_j}{\ln(d)} < \infty$.*

In both cases the ε -exponent is at most 2, which comes from the averaging argument where we used the Gaussian-weighted root-mean-square error. However, we are convinced that there

are quasi-Monte Carlo methods for which the ε -exponent can be improved. The d -exponent is given by $\zeta(\alpha_{\min})A$ with A defined in Theorem 3.14 and so it depends on both the weights and the smoothness parameters.

For the Hermite space with exponentially decaying coefficients we can show not only an upper bound but also a lower bound. With it, we are able to state sufficient and necessary conditions on the weight sequence for polynomial tractability.

Proposition 3.15. *Let $\gamma = (\gamma_1, \dots, \gamma_d) \in (0, \infty)^d$ be a non-increasing sequence of weights and $\omega = (\omega_1, \dots, \omega_d) \in (0, 1)^d$ a sequence of smoothness parameters. There exists a quasi-Monte Carlo method with point set \mathcal{P} such that the worst case error for QMC integration in the Hermite space $\mathcal{H}_{\epsilon_{\omega, \gamma}}$ of functions with exponentially decaying coefficients can be bounded from above by*

$$e_{n,d}(\mathcal{P}, \mathcal{H}_{\epsilon_{\omega, \gamma}}) \leq \frac{1}{\sqrt{n}} e^{\frac{\omega_{\max}}{2(1-\omega_{\max})} \sum_{j=1}^d \gamma_j} \quad (16)$$

with $\omega_{\max} := \max\{\omega_i : i = 1, \dots, d\}$.

Proof. With Corollary 3.11 and the definition of $\epsilon_{\omega, \gamma}$, see (10), we know that there exists a QMC method with

$$e_{n,d}(\mathcal{P}, \mathcal{H}_{\epsilon_{\omega, \gamma}}) \leq \frac{1}{\sqrt{n}} e^{\frac{1}{2} \sum_{j=1}^{\infty} \gamma_j \frac{\omega_j}{1-\omega_j}}.$$

The statement directly follows, because $\omega_j/(1-\omega_j) < \omega_{\max}/(1-\omega_{\max})$ for all $j = 1, \dots, d$. \square

We again get sufficient conditions to the weight sequence to achieve polynomial tractability in $\mathcal{H}_{\epsilon_{\omega, \gamma}}$. We have that multivariate integration is strong polynomial tractable if the weight sequence is summable and the integration problem is polynomial tractable if $\limsup_d \sum_{j=1}^d \gamma_j / \ln(d) < \infty$. However, to show that these conditions are not only sufficient but also necessary, we need a lower bound on the worst case error. For that, we have to choose our weights small enough such that the reproducing kernel $K_{\epsilon_{\omega, \gamma}}$ is positive.

Proposition 3.16. *Let $\gamma = (\gamma_1, \dots, \gamma_d) \in (0, 1)^d$ be a non-increasing sequence of weights and $\omega = (\omega_1, \dots, \omega_d) \in (0, 1)^d$ a sequence of smoothness parameters. For any quasi-Monte Carlo method with point set \mathcal{P} the worst case error for QMC integration in the Hermite space $\mathcal{H}_{\epsilon_{\omega, \gamma}}$ of functions with exponentially decaying coefficients can be bounded from below by*

$$e_{n,d}(\mathcal{P}, \mathcal{H}_{\epsilon_{\omega, \gamma}}) \geq \left(-1 + \frac{\prod_{j=1}^d (1 + \gamma_j c(\omega_j))}{n} \right)^{\frac{1}{2}} \quad (17)$$

with $c(\omega) = \frac{1 - \sqrt{1 - \omega^2}}{\sqrt{1 - \omega^2}}$.

Proof. With (13) we have that

$$e_{n,d}^2(\mathcal{P}, \mathcal{H}_{\epsilon_{\omega, \gamma}}) = -1 + \frac{1}{n^2} \sum_{i,j=1}^n K_{\epsilon_{\omega, \gamma}}(\mathbf{x}_i, \mathbf{x}_j)$$

holds for any QMC method with $\mathcal{P} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. Since all weights γ_j are smaller than 1, we know from (11) that $K_{\epsilon_{\omega}, \gamma}(\mathbf{x}_i, \mathbf{x}_j) > 0$ and so we neglect all terms with $i \neq j$. Hence,

$$e_{n,d}^2(\mathcal{P}, \mathcal{H}_{\epsilon_{\omega}, \gamma}) \geq -1 + \frac{1}{n} \sum_{i=1}^n K_{\epsilon_{\omega}, \gamma}(\mathbf{x}_i, \mathbf{x}_i).$$

For any $i = 1, \dots, d$ we get from (11) that

$$\begin{aligned} K_{\epsilon_{\omega}, \gamma}(\mathbf{x}_i, \mathbf{x}_i) &= \prod_{j=1}^d \left(1 - \gamma_j + \gamma_j \frac{1}{\sqrt{1 - \omega_j^2}} \exp \left(\frac{\omega_j}{1 + \omega_j} x_{i,j}^2 \right) \right) \\ &\geq \prod_{j=1}^d \left(1 - \gamma_j + \gamma_j \frac{1}{\sqrt{1 - \omega_j^2}} \right) \end{aligned}$$

where $x_{i,j}$ is the j -th entry of \mathbf{x}_i . Putting everything together, we get

$$e_{n,d}^2(\mathcal{P}, \mathcal{H}_{\epsilon_{\omega}, \gamma}) \geq -1 + \frac{1}{n} \prod_{j=1}^d \left(1 + \gamma_j \frac{1 - \sqrt{1 - \omega_j^2}}{\sqrt{1 - \omega_j^2}} \right).$$

□

Now we assume that γ is a non-increasing infinite sequence of weights $\gamma_j \in (0, 1)$. Moreover, we suppose that $\omega \in (0, 1)^{\mathbb{N}}$ is a sequence such that $\omega_j \geq \omega_{\min} > 0$ for all $j \in \mathbb{N}$. It follows from Proposition 3.16 that

$$\begin{aligned} n_{\min}(\varepsilon, d; \mathcal{H}_{\epsilon_{\omega}, \gamma}) &\geq \frac{\prod_{j=1}^d (1 + \gamma_j c(\omega_j))}{\varepsilon^2 + 1} \\ &\geq \frac{\prod_{j=1}^d (1 + \gamma_j c(\omega_{\min}))}{\varepsilon^2 + 1} \end{aligned}$$

with $c(\omega) = \frac{1 - \sqrt{1 - \omega^2}}{\sqrt{1 - \omega^2}}$. Under the assumption that $\gamma_j \geq \gamma_{\min} > 0$ for all $j \in \mathbb{N}$, we get

$$n_{\min}(\varepsilon, d; \mathcal{H}_{\epsilon_{\omega}, \gamma}) \geq \frac{(1 + \gamma_{\min} c(\omega_{\min}))^d}{\varepsilon^2 + 1}$$

and thus the information complexity grows exponentially in d . If we assume that $\gamma_j \rightarrow 0$ and $\sum_{j=1}^{\infty} \gamma_j = \infty$, we can deduce from

$$\prod_{j=1}^d (1 + \gamma_j c(\omega_j)) \geq e^{\tilde{c}(\omega_0) \sum_{j=1}^d \gamma_j}$$

with constant $\tilde{c} > 0$, that $n_{\min}(\varepsilon, d; \mathcal{H}_{\epsilon_{\omega}, \gamma}) \rightarrow \infty$ as $d \rightarrow \infty$. Therefore $\sum_{j=1}^{\infty} \gamma_j < \infty$ is a necessary condition for strong polynomial tractability. Next we assume that $\gamma_j \rightarrow 0$ for all

$j \in \mathbb{N}$ and $\limsup_d \sum_{j=1}^d \gamma_j / \ln(d) = \infty$. Then we can use the estimate

$$\prod_{j=1}^d (1 + \gamma_j c(\omega_j)) \geq d^{\bar{c}c(\omega_0) \sum_{j=1}^d \gamma_j / \ln(d)}$$

with constant $\bar{c} > 0$ to conclude that n_{\min} grows faster than any power of d . Thus we can conclude that $\limsup_d \sum_{j=1}^d \gamma_j / \ln(d) < \infty$ is necessary to get polynomial tractability. Now we summarize the results concerning the Hermite space with exponentially decaying coefficients in the following theorem.

Theorem 3.17. *Let $\gamma = (\gamma_1, \gamma_2, \dots) \in (0, 1)^{\mathbb{N}}$ be a non-increasing sequence of weights and $\omega = (\omega_1, \omega_2, \dots) \in (0, 1)^{\mathbb{N}}$ be a sequence of smoothness parameters such that $\liminf_d \omega_d > 0$ and $\limsup_d \omega_d < 1$. Multivariate integration in the weighted Hermite space $\mathcal{H}_{\epsilon_{\omega}, \gamma}$ of functions with exponentially decaying coefficients is*

1. *strong polynomially tractable iff $\sum_{j=1}^{\infty} \gamma_j < \infty$,*
2. *polynomially tractable iff $\limsup_d \frac{\sum_{j=1}^d \gamma_j}{\ln(d)} < \infty$.*

4 Quasi-Monte Carlo integration and orthogonal transforms

In this section we study the influence of orthogonal transforms on the efficiency of QMC algorithm.

4.1 Error analysis

First, let us summarize the results which we have until now. Due to (2) we know that the error of the quasi-Monte Carlo algorithm is bounded by the sum of the discretization error and the integration error. For the discretization error we have the estimate (3) which states that the error can be reduced by increasing the fineness of the discretization grid. The convergence rate of the discretization step depends on the given time-continuous problem as well as on the discretization method. Moreover, we denote by f_d the d -dimensional function which we obtain after the discretization and we assume that for every d the function f_d belongs to a weighted Hermite space in which multivariate integration is polynomially tractable. Then,

$$|\mathbb{E}(g(B)) - Q_{n,d}| \leq c_1 d^{-q_1} + \|f_d\|_{\mathbf{r}} c_2 d^{q_2} n^{-\frac{1}{2}} \quad (18)$$

with constants $c_1, c_2 > 0$. The exponent $q_1 > 0$ corresponds to the discretization method and $q_2 > 0$ is the d -exponent. However, it can happen that the norm of f_d grows in d , i.e., the sequence $(\|f_d\|_{\mathbf{r}})_{d \in \mathbb{N}} \rightarrow \infty$ as $d \rightarrow \infty$. Then, even if we have strong polynomial tractability in the Hermite space, i.e., $q_2 = 0$, the second term still gets bigger with increasing d , sometimes rendering bound (18) useless.

So let us revisit the estimation of the integration error. Because of the change-of-variable formula for multi-dimensional integration we have for any orthogonal transform $U : \mathbb{R}^d \rightarrow \mathbb{R}^d$

and for any $f_d \in L^2(\mathbb{R}^d, \varphi_d)$ the identity

$$\int_{\mathbb{R}^d} f_d(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} f_d(U\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x}.$$

Thus, instead of approximating the integral on the left-hand side we can compute the integral on the right-hand side by a quasi-Monte Carlo rule. Then we can replace the norm of f_d by the norm of $f_d \circ U$ in the upper bound of the integration error in (18). Now the idea is to choose the orthogonal transform U such that $\|f_d \circ U\|_r$ increases slower in d than the norm of f_d does or to even get that $\|f_d \circ U\|_r$ is bounded.

For that, we would like to know how an orthogonal transform affects the norm $\|\cdot\|_r$ of the Hermite space \mathcal{H}_r , but at first we consider an example to show what one can achieve by applying an orthogonal transform to the integration problem.

Example 4.1. We are interested in the computation of $\mathbb{E}(\exp(B_1))$ where $B = (B_t)_{t \in [0,1]}$ is a standard Brownian motion. Note that this problem is of the form of the model problem presented in the introduction. The first attempt is to apply the forward method to discretize the Brownian motion on the time grid $\frac{1}{d}, \frac{2}{d}, \dots, \frac{d}{d}$. Then the discretized problem is given by the d -dimensional function $f_d(\mathbf{x}) = \exp(\frac{1}{\sqrt{d}} \sum_{j=1}^d x_j)$ which belongs to $\mathcal{H}_{\mathbf{p}_{\alpha, \gamma}}$. We further assume that α is a d -dimensional vector of integers larger than 1. For the \mathbf{k} -th Hermite coefficient we get, by using the exponential generating function,

$$\begin{aligned} \widehat{f_d}(\mathbf{k}) &= \int_{\mathbb{R}^d} e^{\frac{1}{\sqrt{d}} \sum_{j=1}^d x_j} H_{\mathbf{k}}(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= \prod_{j=1}^d \int_{\mathbb{R}} e^{\frac{1}{\sqrt{d}} x_j} H_{k_j}(x_j) \varphi(x_j) dx_j \\ &= e^{\frac{1}{2}} \prod_{j=1}^d \int_{\mathbb{R}} \sum_{\ell=0}^{\infty} \frac{1}{\sqrt{\ell! d^\ell}} H_\ell(x_j) H_{k_j}(x_j) \varphi(x_j) dx_j \\ &= e^{\frac{1}{2}} \prod_{j=1}^d \frac{1}{\sqrt{k_j! d^{k_j}}} \\ &= e^{\frac{1}{2}} \frac{1}{\sqrt{\mathbf{k}! d^{|\mathbf{k}|}}}. \end{aligned}$$

The norm of f_d is given by

$$\begin{aligned} \|f_d\|_{\mathbf{p}_{\alpha, \gamma}}^2 &= e \prod_{j=1}^d \left(1 + \gamma_j^{-1} \sum_{k=1}^{\infty} k^{\alpha_j} \frac{1}{k! d^k} \right) \\ &= e \prod_{j=1}^d \left(1 + \gamma_j^{-1} \left(\frac{1}{d} m_{\alpha_j}(1/d) e^{1/d} \right) \right) \\ &= e^{1 + \sum_{j=1}^d \ln(1 + \gamma_j^{-1} (\frac{1}{d} m_{\alpha_j}(1/d) e^{1/d}))} \end{aligned}$$

where m_{α} is a polynomial of degree $\alpha - 1$. Moreover, $m_{\alpha}(1/d)$ is monotonically decreasing in d and $m_{\alpha}(0) = 1$. If we now choose the weights γ as $\gamma_j = j^{-2}$, we would achieve that integration in $\mathcal{H}_{\mathbf{p}_{\alpha, \gamma}}$ is strong polynomially tractable, due to Theorem 3.14. However, we can bound $\|f_d\|_{\mathbf{p}_{\alpha, \gamma}}$

from below,

$$\begin{aligned}\|f_d\|_{\mathbf{p}_{\alpha,\gamma}}^2 &\geq e^{1+\sum_{j=1}^d \ln\left(\frac{j^2}{d}\right)} \\ &= e^{1+(2\ln(d!)-d\ln(d))} \\ &= e^{\frac{(d!)^2}{d^d}}.\end{aligned}$$

Since $\frac{(d!)^2}{d^d} \rightarrow \infty$ as $d \rightarrow \infty$, the sequence of norms $(\|f_d\|_{\mathbf{p}_{\alpha,\gamma}})_{d \in \mathbb{N}}$ grows in d . Thus, even though we have strong polynomial tractability of the integration problem the upper bound of the integration error is not independent of the dimension, because the norm still depends on d .

Now let us revisit the problem: Instead of the forward method we use the Brownian bridge construction method to discretize the Brownian motion on the time grid $\frac{1}{d}, \frac{2}{d}, \dots, \frac{d}{d}$. In other words, we apply an orthogonal transform U , namely the inverse Haar transform, to the integration problem. Then the discretized problem is given by the function $(f_d \circ U)(\mathbf{x}) = \exp(x_1)$ which belongs to the same Hermite space as f_d . The \mathbf{k} -the Hermite coefficient is given by

$$\widehat{f_d \circ U}(\mathbf{k}) = \begin{cases} \sqrt{\frac{e}{k_1!}} & \text{if } k_2 = \dots = k_d = 0 \\ 0 & \text{else} \end{cases}$$

and consequently,

$$\begin{aligned}\|f_d \circ U\|_{\mathbf{p}_{\alpha,\gamma}}^2 &= e + \gamma_1^{-1} \sum_{k=1}^{\infty} k_1^{\alpha_1} \frac{e}{k_1!} \\ &= e \left(1 + \gamma_1^{-1} m_{\alpha_1}(1)e\right).\end{aligned}$$

This means that the upper bound of the integration error does not depend on d if we choose our weights γ such that strong polynomial tractability holds.

4.2 Orthogonal transforms

As we have seen above, the convergence of a quasi-Monte Carlo algorithm can be influenced by applying an orthogonal transform to the integration problem. However, it is crucial to choose the orthogonal transform tailored to the integration problem to improve the efficiency of the quasi-Monte Carlo algorithm, but in general this is not an easy problem at all. A first attempt to get a better understanding of the behavior of the orthogonal transforms is to study the relation of the Hermite coefficients of f and $f \circ U$.

For an orthogonal transform U of the \mathbb{R}^d we define the mapping $\mathcal{A}_U : L^2(\mathbb{R}^d, \varphi_d) \rightarrow L^2(\mathbb{R}^d, \varphi_d)$ as the composition of f and U , i.e., $\mathcal{A}_U f := f \circ U$.

Theorem 4.2. *Let U be an orthogonal transform of the \mathbb{R}^d . The mapping \mathcal{A}_U is a Hilbert space automorphism on $L^2(\mathbb{R}^d, \varphi_d)$ and furthermore $(\mathcal{A}_U)^{-1} = \mathcal{A}_{U^{-1}}$.*

Proof. Let $f, g \in L^2(\mathbb{R}^d, \varphi_d)$. Then

$$\begin{aligned}\langle \mathcal{A}_U f, \mathcal{A}_U g \rangle_{L^2(\mathbb{R}^d, \varphi_d)} &= \int_{\mathbb{R}^d} (\mathcal{A}_U f)(\mathbf{x})(\mathcal{A}_U g)(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} f(U\mathbf{x})g(U\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} f(\mathbf{x})g(\mathbf{x}) \varphi_d(\mathbf{x}) d\mathbf{x} \\ &= \langle f, g \rangle_{L^2(\mathbb{R}^d, \varphi_d)}\end{aligned}$$

and so we have that \mathcal{A}_U preserves the inner product. Moreover, we have that \mathcal{A}_U is linear as well as surjective. Thus, \mathcal{A}_U is a Hilbert space automorphism on $L^2(\mathbb{R}^d, \varphi_d)$. Since for any $\mathbf{x} \in \mathbb{R}^d$

$$\mathcal{A}_U \mathcal{A}_{U^{-1}} f(\mathbf{x}) = f(UU^{-1}\mathbf{x}) = f(U^{-1}U\mathbf{x}) = \mathcal{A}_{U^{-1}} \mathcal{A}_U f(\mathbf{x}),$$

the inverse of \mathcal{A}_U is given by $\mathcal{A}_{U^{-1}}$. □

Apart from this, we look for a nice representation of \mathcal{A}_U . For that, we define sub-spaces of the $L^2(\mathbb{R}^d, \varphi_d)$ as the linear span of the Hermite polynomials of the same degree, i.e., $\mathcal{H}^{(m)} := \text{span}\{H_{\mathbf{k}}(\mathbf{x}) : |\mathbf{k}| = m\}$ for any $m \in \mathbb{N}_0$. Since the Hermite polynomials form a basis of the $L^2(\mathbb{R}^d, \varphi_d)$, it is easy to deduce that

$$L^2(\mathbb{R}^d, \varphi_d) = \overline{\bigoplus_{m=0}^{\infty} \mathcal{H}^{(m)}}.$$

Recall that the \mathbf{k} -th Hermite polynomial is given by

$$H_{\mathbf{k}}(\mathbf{x}) = \frac{\partial^{|\mathbf{k}|}}{\partial \mathbf{t}^{\mathbf{k}}} G(\mathbf{x}, \mathbf{t}) \Big|_{\mathbf{t}=\mathbf{0}}$$

where G is the exponential generator function of the Hermite polynomials. For any orthogonal transform U of the \mathbb{R}^d we know that

$$\begin{aligned}G(U\mathbf{x}, \mathbf{t}) &= e^{(U\mathbf{x})^\top \mathbf{t} - \frac{\mathbf{t}^\top \mathbf{t}}{2}} \\ &= e^{\mathbf{x}^\top (U^\top \mathbf{t}) - \frac{(U^\top \mathbf{t})^\top (U^\top \mathbf{t})}{2}} \\ &= G(\mathbf{x}, U^\top \mathbf{t})\end{aligned}$$

holds, and so we obtain

$$\begin{aligned}\frac{\partial}{\partial t_j} G(U\mathbf{x}, \mathbf{t}) \Big|_{\mathbf{t}=\mathbf{0}} &= \frac{\partial}{\partial t_j} G(\mathbf{x}, U^\top \mathbf{t}) \Big|_{\mathbf{t}=\mathbf{0}} \\ &= \sum_{k=1}^d U_{jk} \left(\frac{\partial}{\partial t_k} G(\mathbf{x}, \mathbf{t}) \Big|_{\mathbf{t}=\mathbf{0}} \right).\end{aligned}$$

In the same way, we have for a sequence $(\beta_1, \dots, \beta_m) \in \{1, \dots, d\}^m$ of indices

$$\frac{\partial}{\partial t_{\beta_m}} \cdots \frac{\partial}{\partial t_{\beta_1}} G(U\mathbf{x}, \mathbf{t}) \Big|_{\mathbf{t}=\mathbf{0}} = \sum_{\xi_1, \dots, \xi_m=1}^d U_{\beta_m \xi_m} \cdots U_{\beta_1 \xi_1} \left(\frac{\partial}{\partial t_{\xi_m}} \cdots \frac{\partial}{\partial t_{\xi_1}} G(\mathbf{x}, \mathbf{t}) \Big|_{\mathbf{t}=\mathbf{0}} \right). \quad (19)$$

Since there are only derivatives of order m involved, $H_{\mathbf{k}}(U\mathbf{x})$ with $|\mathbf{k}| = m$ can be written as a linear combination of Hermite polynomials of degree m . Therefore, we can state the following lemma.

Lemma 4.3. *The restriction of \mathcal{A}_U to $\mathcal{H}^{(m)}$, denoted by $\mathcal{A}_{U,m}$, is a Hilbert space automorphism of $\mathcal{H}^{(m)}$.*

Note that the sum in (19) does not represent the left-hand side in a unique way, because there are many $\beta \in \{1, \dots, d\}^m$ which correspond to the same multi-index. So we consider the function $S : \{1, \dots, d\}^m \rightarrow \mathbb{N}_0^d$, given by

$$S((\beta_1, \dots, \beta_m)) := (\#\{k : \beta_k = 1\}, \dots, \#\{k : \beta_k = d\}),$$

which gives for each sequence $(\beta_1, \dots, \beta_m)$ the corresponding multi-index.

Lemma 4.4. *Let $\mathbf{k} \in \mathbb{N}_0^d$ be a multi-index with $|\mathbf{k}| = m$. Then*

$$\#\{\beta \in \{1, \dots, d\}^m : S(\beta) = \mathbf{k}\} = \frac{|\mathbf{k}|!}{\mathbf{k}!}.$$

Proof. This can be seen by noting that the right-hand side is just a multinomial coefficient, $\frac{|\mathbf{k}|!}{\mathbf{k}!} = \binom{k_1 + \dots + k_m}{k_1, \dots, k_m}$, and by using elementary combinatorics. \square

To use the representation of equation (19) we want a space that takes account of the order of differentiation,

$$\mathcal{K}^{(m)} := \ell^2(\{1, \dots, d\}^m) = \{w : \{1, \dots, d\}^m \rightarrow \mathbb{R}\}$$

with norm

$$\|w\|_{\ell^2} = \sum_{\beta \in \{1, \dots, d\}^m} w_{\beta}^2.$$

Denote the canonical basis of $\mathcal{K}^{(m)}$ by $(b_{\beta})_{\beta \in \{1, \dots, d\}^m}$. Next we define the linear operator,

$$\begin{aligned} \mathcal{J}_m : \mathcal{H}^{(m)} &\longrightarrow \mathcal{K}^{(m)} \\ H_{\mathbf{k}} &\longmapsto \sqrt{\frac{\mathbf{k}!}{|\mathbf{k}|!}} \sum_{S(\beta)=\mathbf{k}} b_{\beta}, \end{aligned}$$

the adjoint operator of which is given by

$$\mathcal{J}_m^* : \mathcal{K}^{(m)} \longrightarrow \mathcal{H}^{(m)}$$

$$b_\beta \mapsto \sqrt{\frac{S(\beta)!}{|S(\beta)|!}} H_{S(\beta)}.$$

Proposition 4.5. *The linear operator \mathcal{J}_m is an isometry and consequently, $\mathcal{J}_m^* \mathcal{J}_m = \text{id}_{\mathcal{H}^m}$*

Proof. Let $f \in \mathcal{H}^{(m)}$ with $f(\mathbf{x}) = \sum_{|\mathbf{k}|=m} f_{\mathbf{k}} H_{\mathbf{k}}(\mathbf{x})$. Then,

$$\|f\|_{\mathcal{H}^{(m)}}^2 = \sum_{|\mathbf{k}|=m} f_{\mathbf{k}}^2$$

and

$$\begin{aligned} \mathcal{J}_m f &= \sum_{|\mathbf{k}|=m} f_{\mathbf{k}} \mathcal{J}_m H_{\mathbf{k}} = \sum_{|\mathbf{k}|=m} f_{\mathbf{k}} \sqrt{\frac{\mathbf{k}!}{|\mathbf{k}|!}} \sum_{S(\beta)=\mathbf{k}} b_\beta \\ &= \sum_{\beta \in \{1, \dots, d\}^m} \sum_{\mathbf{k}=S(\beta)} f_{\mathbf{k}} \sqrt{\frac{\mathbf{k}!}{|\mathbf{k}|!}} b_\beta \\ &= \sum_{\beta \in \{1, \dots, d\}^m} f_{S(\beta)} \sqrt{\frac{S(\beta)!}{|S(\beta)|!}} b_\beta. \end{aligned}$$

So we have

$$\begin{aligned} \|\mathcal{J}_m f\|_{\mathcal{K}^{(m)}} &= \sum_{\beta \in \{1, \dots, d\}^m} f_{S(\beta)}^2 \frac{S(\beta)!}{|S(\beta)|!} \\ &= \sum_{|\mathbf{k}|=m} \sum_{S(\beta)=\mathbf{k}} f_{S(\beta)}^2 \frac{S(\beta)!}{|S(\beta)|!} \\ &= \sum_{|\mathbf{k}|=m} f_{\mathbf{k}}^2 \sum_{S(\beta)=\mathbf{k}} \frac{S(\beta)!}{|S(\beta)|!}, \end{aligned}$$

With Lemma 4.4 we get that $\sum_{S(\beta)=\mathbf{k}} \frac{S(\beta)!}{|S(\beta)|!} = 1$ and consequently, $\|f\|_{\mathcal{H}^{(m)}} = \|\mathcal{J}_m f\|_{\mathcal{K}^{(m)}}$. \square

According to equation (19) the application of an orthogonal transform U can be represented in the space $\mathcal{K}^{(m)}$ as a matrix-vector multiplication with the matrix given by the m -fold Kronecker product of U :

$$U^{\otimes m} := \underbrace{U \otimes \dots \otimes U}_{m \text{ times}}.$$

Therefore, the mapping $\mathcal{A}_{U,m}$ is of the form

$$\mathcal{A}_{U,m} = \mathcal{J}_m^* U^{\otimes m} \mathcal{J}_m$$

which can be illustrated by the commutative diagram:

$$\begin{array}{ccc}
\mathcal{H}^{(m)} & \xrightarrow{\mathcal{A}_{U,m}} & \mathcal{H}^{(m)} \\
\mathcal{J}_m \downarrow & & \uparrow \mathcal{J}_m^* \\
\mathcal{K}^{(m)} & \xrightarrow{U^{\otimes m}} & \mathcal{K}^{(m)}
\end{array}$$

So we have a nice representation of $\mathcal{A}_{U,m}$ on the space spanned by the Hermite polynomials of degree m and we can deduce a similar representation of \mathcal{A}_U on $L^2(\mathbb{R}^d, \varphi_d)$. This can be achieved by using

$$\mathcal{A}_U = \bigoplus_{m=0}^{\infty} \mathcal{A}_{U,m}.$$

We define the space $\mathcal{K} := \overline{\bigoplus_{m=0}^{\infty} \mathcal{K}^{(m)}}$ and the isometry $\mathcal{J} : L^2(\mathbb{R}^d, \varphi_d) \longrightarrow \mathcal{K}$ by

$$\mathcal{J} := \bigoplus_{m=0}^{\infty} \mathcal{J}_m,$$

the adjoint operator of which is given by

$$\mathcal{J}^* := \bigoplus_{m=0}^{\infty} \mathcal{J}_m^*.$$

This means that we can represent the mapping \mathcal{A}_U by

$$\mathcal{A}_U = \mathcal{J}^* \bigoplus_{m=0}^{\infty} U^{\otimes m} \mathcal{J} \tag{20}$$

and we have the commutative diagram:

$$\begin{array}{ccc}
L^2(\mathbb{R}^d, \varphi_d) & \xrightarrow{\mathcal{A}_U} & L^2(\mathbb{R}^d, \varphi_d) \\
\mathcal{J} \downarrow & & \uparrow \mathcal{J}^* \\
\mathcal{K} & \xrightarrow{\bigoplus_{m \geq 0} U^{\otimes m}} & \mathcal{K}
\end{array}$$

In addition to being an elegant representation of \mathcal{A}_U , equation (20) provides us with a formula for computing the Hermite coefficients of $f \circ U$ from those of f . For a better illustration we consider an example.

Example 4.6. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ with Hermite expansion $f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^2} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x})$ and U be an orthogonal transform with matrix representation

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}.$$

Since we know that \mathcal{A}_U factors nicely with respect to the spaces $\mathcal{H}^{(m)}$, we restrict us to the case

$m = 2$ and consider $\mathcal{A}_{U,2}$ as well as the projection f_2 of f onto the space $\mathcal{H}^{(2)}$, i.e., $f_2(\mathbf{x}) = f(\mathbf{x})|_{\mathcal{H}^{(2)}} = \sum_{|\mathbf{k}|=2} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x})$. Thus f_2 is fully given by the vector of Hermite coefficients $(\hat{f}(2,0), \hat{f}(1,1), \hat{f}(0,2))^\top$. Now we get the Hermite coefficients of $\mathcal{A}_{U,2}f_2$ by standard matrix-vector calculus. The operator \mathcal{J}_2 , mapping from $\mathcal{H}^{(2)}$ to $\mathcal{K}^{(2)}$, has the matrix representation

$$\mathcal{J}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and \mathcal{J}_2^* is given by the transpose of \mathcal{J}_2 . Moreover, $U^{\otimes 2}$ can also be written as a matrix. Thus, we get the vector containing the Hermite coefficients of $\mathcal{A}_{U,2}f_2$ by matrix-vector multiplication, i.e.,

$$\left(\widehat{\mathcal{A}_{U,2}f_2}(2,0), \widehat{\mathcal{A}_{U,2}f_2}(1,1), \widehat{\mathcal{A}_{U,2}f_2}(0,2) \right)^\top = \mathcal{J}_2^\top U^{\otimes 2} \mathcal{J}_2 \left(\hat{f}(2,0), \hat{f}(1,1), \hat{f}(0,2) \right)^\top.$$

Since it suffices to know the Hermite coefficients of a function, we fully know how the orthogonal transform affects the function f_2 . \square

Because of Theorem 4.2 we have that the norm of any function $f \in L^2(\mathbb{R}^d, \varphi_d)$ can not be influenced by applying a mapping \mathcal{A}_U , but we are interested in functions f which belong to a Hermite space \mathcal{H}_r with norm $\|\cdot\|_r$. To preserve the representation (20) of \mathcal{A}_U , we have to equip \mathcal{K} with a suitable norm. For that, we introduce, for any $m \in \mathbb{N}_0$, on the space $\mathcal{K}^{(m)}$ the norm

$$\|w\|_{\ell^2, r} = \sum_{\beta \in \{1, \dots, d\}^m} r(S(\beta))^{-1} w_\beta^2. \quad (21)$$

Then it follows that \mathcal{J} , as defined above, is still a isometry between \mathcal{H}_r and \mathcal{K} as well as $\mathcal{J}^* \mathcal{J} = \text{id}_{\mathcal{H}_r}$. Nevertheless, $\bigoplus_{m=0}^\infty U^{\otimes m}$ is not a Hilbert space automorphism on \mathcal{K} equipped with the norm (21). Consequently, we get that, in general,

$$\|f\|_r \neq \|\mathcal{A}_U f\|_r.$$

Thus, we can use orthogonal transforms in the sense that we choose for a given function $f \in \mathcal{H}_r$ an orthogonal transform U such that the norm of $\mathcal{A}_U f$ grows slower in d than $\|f\|_r$.

The problem of finding an orthogonal transform that makes the norm smaller is not an easy one. However, note that there exists at least one orthogonal transform such that the transformed problem is not worse than the original one, namely the identity.

In Irrgeher and Leobacher [10, 11] a “regression algorithm” is introduced which determines an orthogonal transform tailored to the given problem which is optimal in some sense. The idea of the algorithm is to approximate the function f by a linear function using a linear regression approach. For the linear function it is easy to determine an orthogonal transform so that the linear function becomes one-dimensional. Then this orthogonal transform is used together with a QMC rule to solve the original integration problem.

It turns out that this algorithm fits well into the setting of Hermite spaces. Let f be in a

Hermite space \mathcal{H}_r . Because of Theorem 3.2 we know that f has a pointwise convergent Hermite expansion, i.e.,

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^d$. The first step of the regression algorithm is the linear regression approach and it is easy to see that this means that we approximate the function f by using the truncated Hermite expansion of order 1, i.e.,

$$f(\mathbf{x}) \approx \sum_{|\mathbf{k}| \leq 1} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(\mathbf{x}).$$

We denote the multi-index $(0, \dots, 1, \dots, 0)$ with 1 at the j -th entry by \mathbf{e}_j and we write $\mathbf{b} = (\hat{f}(\mathbf{e}_1), \dots, \hat{f}(\mathbf{e}_d))$. Now we determine an orthogonal transform U , namely a Householder reflection, which maps the vector $(1, 0, \dots, 0)^\top$ to $\mathbf{b}/\|\mathbf{b}\|_{\ell_2}$, where $\|\cdot\|_{\ell_2}$ denotes the euclidean norm and where we assume that the linear part \mathbf{b} of f does not vanish. So we have that

$$f(U\mathbf{x}) = \hat{f}(\mathbf{0}) + \|\mathbf{b}\|_{\ell_2} x_1 + \sum_{|\mathbf{k}| \geq 2} \hat{f}(\mathbf{k}) H_{\mathbf{k}}(U\mathbf{x}),$$

and we see that linear part of the Hermite expansion depends only on x_1 . If we now consider the norm of $f \circ U$ we get that

$$\|f \circ U\|_r^2 = \hat{f}(\mathbf{0})^2 + r(\mathbf{e}_1)^{-1} \|\mathbf{b}\|_{\ell_2}^2 + \sum_{|\mathbf{k}| \geq 2} r(\mathbf{k})^{-1} \widehat{f \circ U}(\mathbf{k})^2.$$

If f is such that its linear part already covers almost all of the \mathcal{H}_r -norm, it may happen that U reduces significantly the norm, which, in fact, depends on the function r . For example, let us consider $r = p_{\alpha, \gamma}$ with $\gamma_1 > \gamma_2 \geq \gamma_3 \geq \dots$. Then,

$$\|f \circ U\|_{p_{\alpha, \gamma}}^2 = \hat{f}(\mathbf{0})^2 + \gamma_1^{-1} \|\mathbf{b}\|_{\ell_2}^2 + \sum_{|\mathbf{k}| \geq 2} p_{\alpha, \gamma}^{-1}(\mathbf{k}) \widehat{f \circ U}(\mathbf{k})^2$$

and

$$\gamma_1^{-1} \|\mathbf{b}\|_{\ell_2}^2 = \gamma_1^{-1} \sum_{j=1}^d \hat{f}(\mathbf{e}_j)^2 < \sum_{j=1}^d \gamma_j^{-1} \hat{f}(\mathbf{e}_j)^2 = \sum_{|\mathbf{k}|=1} p_{\alpha, \gamma}^{-1}(\mathbf{k}) |\hat{f}(\mathbf{k})|^2.$$

Thus, we see that the part of the norm coming from the linear part of f can be reduced by applying the orthogonal transform U . That means that, under certain conditions on the influence of the linear fraction of the function as well as on the weighted Hermite space, the regression algorithm can strictly reduce the norm of the transformed function.

In Irrgeher and Leobacher [10] realistic examples coming from finance are considered. Numerical results show that orthogonal transforms, and in particular the ones computed by the regression algorithm, also lead to more efficient QMC integration.

Furthermore, we want to mention that it can happen that a function does not belong to the Hermite space \mathcal{H}_r , but after applying an orthogonal transform the function is in \mathcal{H}_r . To illustrate that we consider the following simple example.

Example 4.7. Let $\alpha_1 > \alpha_2 > 1$ and $\gamma_1, \gamma_2 > 0$. Now we consider function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ of the form $f(\mathbf{x}) = f_1(x_1)f_2(x_2)$ with continuous and Gaussian square-integrable functions $f_1, f_2 : \mathbb{R} \rightarrow \mathbb{R}$. Furthermore, we suppose that f_1 is such that $\sum_{k=1}^{\infty} k^{\alpha_1} |\hat{f}_1(k)|^2 = \infty$ but $\sum_{k=1}^{\infty} k^{\alpha_2} |\hat{f}_1(k)|^2 < \infty$. This means that the Hermite coefficients of f_1 decay fast enough such that $\|f_1\|_{p_{\alpha_2, \gamma_2}} < \infty$, but for the smoothness parameter α_1 the decay is too slow, i.e. $\|f_1\|_{p_{\alpha_1, \gamma_1}} = \infty$. For the function f_2 we assume that $0 < \sum_{k=1}^{\infty} k^{\alpha_i} |\hat{f}_2(k)|^2 < \infty$ holds for $i = 1, 2$. Then, f does not belong to the Hermite space $\mathcal{H}_{p_{(\alpha_1, \alpha_2), (\gamma_1, \gamma_2)}}$, because

$$\|f\|_{p_{(\alpha_1, \alpha_2), (\gamma_1, \gamma_2)}} = \|f_1\|_{p_{\alpha_1, \gamma_1}} \|f_2\|_{p_{\alpha_2, \gamma_2}} = \infty.$$

However, let us consider the orthogonal transform U , given by

$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Applying the transform to the function f we have $\mathcal{A}_U f(\mathbf{x}) = f(U\mathbf{x}) = f_1(x_2)f_2(x_1)$. Hence,

$$\|\mathcal{A}_U f\|_{p_{(\alpha_1, \alpha_2), (\gamma_1, \gamma_2)}} = \|f_1\|_{p_{\alpha_2, \gamma_2}} \|f_2\|_{p_{\alpha_1, \gamma_1}} < \infty$$

and $\mathcal{A}_U f \in \mathcal{H}_{p_{(\alpha_1, \alpha_2), (\gamma_1, \gamma_2)}}$. Although f itself is not in the Hermite space, we know that the error bound (18) holds if we apply the orthogonal transform U to the function f .

5 Conclusion

We have studied the effect of orthogonal transforms on the effectivity of QMC integration. We have put forward rigorous models in the framework of weighted norms and reproducing kernel Hilbert spaces. Furthermore, we have shown how to compute the norm of a function that is concatenated with an orthogonal transform. This makes it possible to measure whether the orthogonal transform makes the weighted norm smaller.

Of course, it would be desirable to have an algorithm at hand for finding an optimal (or at least a good) orthogonal transform for a given function that minimizes the weighted norm of the transformed function. This has already been tried in Imai and Tan [9] as well as in Irrgeher and Leobacher [10, 11], but those approaches rely on linear approximations of the function and therefore may fail spectacularly for simple but nonlinear functions. We have paved the way for more general algorithms that are similar in spirit but take higher orders of the Hermite expansion into account. But a practical algorithm that, for example, proceeds by approximating the integrand by a quadratic polynomial, has yet to be developed.

We have defined and studied multivariate integration in weighted Hermite spaces. We have succeeded in giving sufficient conditions on the weight sequence for polynomial tractability and strong polynomial tractability in weighted Hermite spaces with polynomially decaying coeffi-

cients. Necessary conditions have yet to be found, but we conjecture that they might coincide with the sufficient ones. Moreover, we have given necessary and sufficient conditions on polynomial tractability and strong polynomial tractability for a class of weighted Hermite spaces with exponentially decaying coefficients.

The question of finding concrete point sets for QMC integration in the spaces considered is completely open. Likely candidates are maps of classical low-discrepancy sequences in the unit hypercube to the \mathbb{R}^d via the inverse cumulative distribution function of the standard normal distribution.

In any case we have contributed to the problem of explaining the efficiency of QMC integration for very high-dimensional problems by providing a general and flexible framework, namely that of Hermite spaces and orthogonal transforms, and by giving important first results on the subject in this framework.

Acknowledgments

The authors are supported by the Austrian Science Fund (FWF): Projects F5508-N26 (Leobacher) and F5509-N26 (Irrgeher), respectively. The projects F5508-N26 and F5509-N26 are part of the Special Research Program “Quasi-Monte Carlo Methods: Theory and Applications”. The first author is also partially supported by the Austrian Science Foundation (FWF), Project P21943.

References

- [1] P. Acworth, M. Broadie, and P. Glasserman. A comparison of some Monte Carlo and quasi-Monte Carlo techniques for option pricing. In H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1996, Proceedings of a Conference at the University of Salzburg, Austria, July 9–12, 1996*, pages 1–18, New York, 1998. Springer.
- [2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.
- [3] T. Björk. *Arbitrage theory in continuous time*. Oxford University Press, third edition, 2009.
- [4] V. I. Bogachev. *Gaussian Measures*. American Mathematical Society, 1998.
- [5] R. E. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance*, 1(1):27–46, 1997.
- [6] H. Cramer. Valeurs Maxima des Polynômes d’Hermite. in Charlier, C. V. L., *Les applications de la théorie des probabilités aux sciences mathématiques et aux sciences physiques; Application de la théorie des probabilités à l’astronomie*. Paris. 1930.
- [7] J. Dick and F. Pillichshammer. *Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration*. Cambridge University Press, 2010.

- [8] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.
- [9] J. Imai and K. S. Tan. A general dimension reduction technique for derivative pricing. *Journal of Computational Finance*, 10:129–155, 2007.
- [10] C. Irrgeher and G. Leobacher. Fast orthogonal transforms for pricing derivatives with quasi-Monte Carlo. In C. Laroque, J. Himmelspace, R. Pasupathy, O. Rose, and A. M. Uhrmacher, editors, *Proceedings of the 2012 Winter Simulation Conference*, 2012.
- [11] C. Irrgeher and G. Leobacher. Fast orthogonal transforms for multilevel quasi-Monte Carlo simulation in computational finance. In Vanmaele, W., et al., editor, *Proceedings of Actuarial and Financial Mathematics Conference 2013: Interplay between Finance and Insurance*, 2013.
- [12] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag Berlin, 2010.
- [13] G. Leobacher. Fast orthogonal transforms and generation of Brownian paths. *Journal of Complexity*, 28:278–302, 2012.
- [14] B. Moskowitz and R. E. Caflisch. Smoothness and dimension reduction in Quasi-Monte Carlo methods. *Mathematical and Computer Modelling*, 23(8-9):37 – 54, 1996.
- [15] A. Papageorgiou. The Brownian bridge does not offer a consistent advantage in quasi-Monte Carlo integration. *Journal of Complexity*, 18(1):171–186, 2002.
- [16] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, Inc., third edition, 1976.
- [17] G. Sansone. *Orthogonal Functions*. R.E. Krieger Publishing Company, 1977.
- [18] I. H. Sloan and H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, 14:1–33, 1998.
- [19] G. Szegő. *Orthogonal polynomials*. American Mathematical Society, fourth edition, 1975.
- [20] S. Thangavelu. *Lectures on Hermite and Laguerre polynomials*. Princeton University Press, 1993.
- [21] X. Wang and I. H. Sloan. Quasi-Monte Carlo Methods in Financial Engineering: An Equivalence Principle and Dimension Reduction. *Operations Research*, 59(1):80–95, 2011.